



CUADERNOS DE TRABAJO

FACULTAD DE ESTUDIOS ESTADÍSTICOS

Análisis Estadístico de las Consultas a la Base de Datos de Ayudas e Incentivos a Empresas de la Dirección General de Industria y de la PYME.

**Adolfo Coello de Portugal Muñoz
Juana María Alonso Revenga**

Cuaderno de Trabajo número 03/2014



UCM

**UNIVERSIDAD
COMPLUTENSE
MADRID**

Los Cuadernos de Trabajo de la Facultad de Estudios Estadísticos constituyen una apuesta por la publicación de los trabajos en curso y de los informes técnicos desarrollados desde la Facultad para servir de apoyo tanto a la docencia como a la investigación.

Los Cuadernos de Trabajo se pueden descargar de la página de la Biblioteca de la Facultad www.ucm.es/BUCM/est/ y en la sección de investigación de la página del centro www.ucm.es/centros/webs/eest/

CONTACTO:

Biblioteca de la Facultad de Estudios Estadísticos

Universidad Complutense de Madrid

Av. Puerta de Hierro, S/N

28040 Madrid

Tlf. 913944035

buc_est@buc.ucm.es

Los trabajos publicados en la serie Cuadernos de Trabajo de la Facultad de Estudios Estadísticos no están sujetos a ninguna evaluación previa. Las opiniones y análisis que aparecen publicados en los Cuadernos de Trabajo son responsabilidad exclusiva de sus autores.

Análisis Estadístico de las Consultas a la Base de Datos de Ayudas e Incentivos a Empresas de la Dirección General de Industria y de la PYME.

Autor: Adolfo Coello de Portugal Muñoz

Coautor: Juana María Alonso Revenga

El Real Decreto 344/2012, de 10 de febrero, publicado el 11 del mismo mes, desarrolla la estructura orgánica básica del Ministerio de Industria, Energía y Turismo, quedando plasmado en dicho real decreto las competencias, actividades y atribuciones de la Dirección General de Industria y de la PYME (en adelante DGIPYME). Entre otras cosas, su articulado le atribuye las siguientes funciones:

Artículo 9, apartado 1, letra g) *La propuesta y ejecución de actuaciones y programas dirigidos a mejorar la competitividad de la PYME, a facilitar su acceso a la financiación y a fomentar el espíritu emprendedor.*

Artículo 10, apartado 1, letra n) *El análisis, las propuestas de mejora, desarrollo y la divulgación de las fuentes de financiación para las pequeñas y medianas empresas. Así como el desarrollo de actuaciones y programas dirigidos a facilitar el acceso a la financiación de la PYME, bien directamente, o instrumentados a través de las sociedades CERSA y ENISA.*

Para poder cumplir con el mandato que obliga a facilitar y divulgar el acceso a la financiación destinadas a las PYME y emprendedores, la DGIPYME ha desarrollado y mantiene, entre otras actuaciones, la Base de Datos de Ayudas e Incentivos para Empresas.

La Base de Datos de Ayudas e Incentivos para Empresas es un servicio de consulta, vía web, a un almacén de datos que recoge las ayudas e incentivos destinados a empresas, otorgados y convocados por la Administración General del Estado, las Administraciones Autonómicas, las Administraciones Locales y otros organismos públicos.

La gestión y el mantenimiento están asignados a un equipo de documentalistas. Diariamente revisan el Boletín Oficial del Estado, los boletines autonómicos y los boletines provinciales en busca de convocatorias de subvenciones, créditos, ayudas e incentivos económicos cuyos beneficiarios son las empresas; incluyendo además la financiación ofertada por diversos organismos públicos.

Una vez revisadas dichas fuentes se recogen todas las convocatorias sobre financiación pública cuyos beneficiarios son empresas o emprendedores. Las convocatorias se estudian y analizan, extractando toda la información necesaria para que la empresa solicite la ayuda. Presentándose ante el ciudadano como un documento pdf al que se asocian los metadatos previamente definidos por el equipo de documentalistas. Los metadatos son el conjunto de conceptos que definen la ayuda y que posibilitarán su búsqueda. Para poder ofrecer al ciudadano el total de la información, se almacena en la base de datos el documento completo de la convocatoria de la ayuda, que también puede ser consultado si se desea.

Resumiendo podemos afirmar que este servicio pone a disposición de quien consulta la siguiente información:

Contenido de la B.D. de Ayudas e Incentivos para Empresas.
<ul style="list-style-type: none"> • Referencia • Título de la ayuda • Organismo convocante • Sector y Subsector al que va dirigida • Administración • Tipo de ayuda y destinatarios • Vigencia y plazo de solicitud • Disposiciones legales reguladoras • Texto íntegro de la disposición

Tabla 1 Relación de contenidos de la B.D. de Ayudas

Con objeto de satisfacer las distintas necesidades y requisitos de información existen varias formas de consulta de las ayudas.

Tipos de consultas a la B.D.
<ul style="list-style-type: none"> • Consulta rápida. • Consulta avanzada. • Consultas prediseñadas. • Registro de consultas en su perfil. • Suscripción por correo electrónico de sus consultas. • Guías dinámicas

Tabla 2. Relación de tipos de consulta a la B.D. de Ayudas.

Las fuentes de financiación son siempre un tema de especial interés, tanto para las empresas como para los emprendedores. La actual crisis ha hecho cesar la actividad de muchas empresas por falta de financiación. Por ello, hoy más que nunca, poner a disposición de las empresas una

información detallada y actualizada que les permita acceder a la financiación pública es una labor necesaria y de gran interés para quienes necesitan liquidez. Sin embargo, los funcionarios de la DGIPYME observaron el escaso número de consultas. Carentes de dinero para costearse una campaña publicitaria que informase sobre los servicios que prestaban, optaron por formas alternativas y gratuitas para difundir sus servicios y darse a conocer.

Jueves, 5 de Septiembre de 2013 - Actualizado a las 23:00 h.

CincoDías

Últimas noticias
09:16 Las lecciones de la crisis
11:19 Madrid 2020: El impacto de
Olimpicos
11:32 Telefónica ultima el lanza
07:49 Almunia: El duopolio es in

[Inicio](#) | [Mercados](#) | [Empresas](#) | [Economía](#) | [Finanzas personales](#) | **[Emprendedores](#)** | [Tecnología](#) | [Multimedia](#) | [Guías](#) | [Ayudas y subvenciones](#) | [Preguntas y respuestas](#) | [Zona Startup](#)

Descarga gratis la única revista de Trading hecha para Traders

Subvenciones para emprendedores

Sector

-Seleccionar uno o varios-

Agrario
Agroalimentario
Agua

Ámbito de aplicación

-Seleccionar uno o varios-

Andalucía
Aragón
Asturias

☐ Ayudas con plazo de solicitud abierto

BUSCAR **BORRAR**

powered by: 

Figura 1 Web del periódico Cinco Días con el Buscador de Ayudas.

Para difundir los servicios que prestan comenzaron a darse de alta en las redes sociales creando sus propios perfiles. Pero lo que les dio más fama y relieve fueron los periódicos económicos digitales. Se dirigieron a ellos para que incluyesen en sus ediciones digitales el buscador de la Base de Datos de Ayudas e Incentivos, ofreciendo este servicio a los diarios y sus lectores de

forma gratuita sin esperar a cambio otra recompensa que difundir una información útil para la sociedad.

The screenshot shows the Expansión.com website interface. At the top, there's a red header with the logo and navigation tabs like 'Mi dinero', 'Empresas', 'Economía', etc. Below this, a green bar displays stock market indices: IBEX 35, I.G. BOLSA MADRID, and DOW JONES. The main content area features a news article titled 'Aprende de los mejores para trazar tu carrera profesional' with a photo of a man. To the right, there's a sidebar titled 'Ayudas e Incentivos' which includes a search form with dropdown menus for 'Sector' (listing Agrario, Agroalimentario, Agua) and 'Ámbito de aplicación' (listing Andalucía, Aragón, Asturias), along with 'Borrar' and 'Buscar' buttons. Below the article, there's another section titled 'Encontrar empleo será más fácil en el próximo trimestre' with a video player and text about job market trends.

Figura 2 Web del periódico Expansión con el Buscador de Ayudas

El resultado de estas colaboraciones fue espectacular. El incremento de las consultas fue altísimo. Se pasó de registrar un mínimo de 11 al día hasta superar las 14.000. Todo ello en un período de dos años y teniendo en cuenta que el público objetivo no es muy numeroso.

El objeto y el interés de este trabajo radica en estudiar y explicar el comportamiento de los datos que recogen el auge de las consultas. Sin olvidar que el estudio y el análisis de los datos no es un fin en sí mismo, sino el camino para poder obtener la información que facilite la toma de decisiones adecuadas.

Análisis y exploración de los datos.

Los datos presentes en esta serie han sido extraídos de los ficheros log del servidor que aloja la base de datos. La información almacenada y recuperada mediante la ejecución de procesos informáticos garantiza la eliminación de errores humanos en el conjunto de datos.

La serie representa el número de accesos diarios a la Base de Datos de Ayudas e Incentivos para Empresas; comenzando el 1 de enero de 2011 y terminando el 11 de marzo de 2013, lo que supone algo más de dos años. Los valores de los accesos obtenidos día a día quedan representados en el siguiente gráfico.

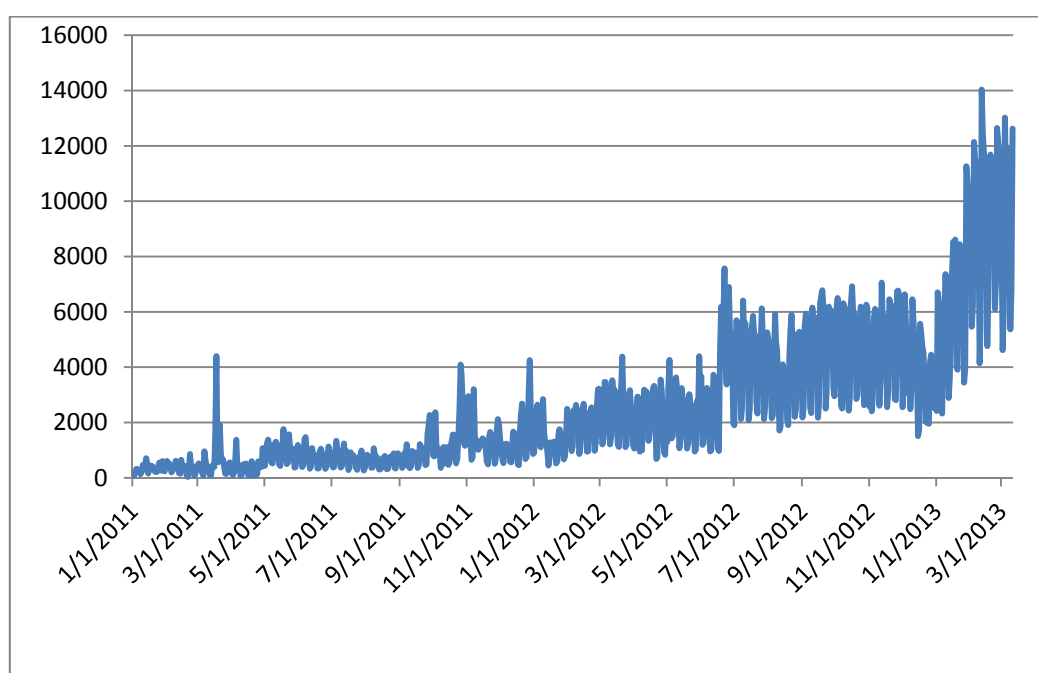


Figura 3 Representación de la serie de las consultas a la B.D. de Ayudas

En un primer análisis de la gráfica se observa como los valores registrados aumentan con el tiempo, partiendo de valores de 11 accesos y llegando a alcanzar los 14.031 accesos por día. Asimismo es fácilmente observable un aumento en la dispersión de los datos, lo que permite afirmar que la serie carece de varianza constante, aumentando ésta con el tiempo.

Llama la atención la presencia de picos. Especialmente uno que aparece al comienzo de la serie situado en el mes de marzo de 2011.

La representación global de toda la serie muestra una gráfica muy abigarrada o densa que impide observar con detalle las oscilaciones y el comportamiento de los datos de la serie. Para una observación más detallada, aunque no general, vamos a representar el mismo mes de la serie en

dos años consecutivos. En concreto, en octubre de 2011 y 2012. La elección del mes de octubre obedece al conocimiento que se tiene de los datos, ya que es muy habitual que una vez finalizado el período vacacional se reanude la actividad económica.

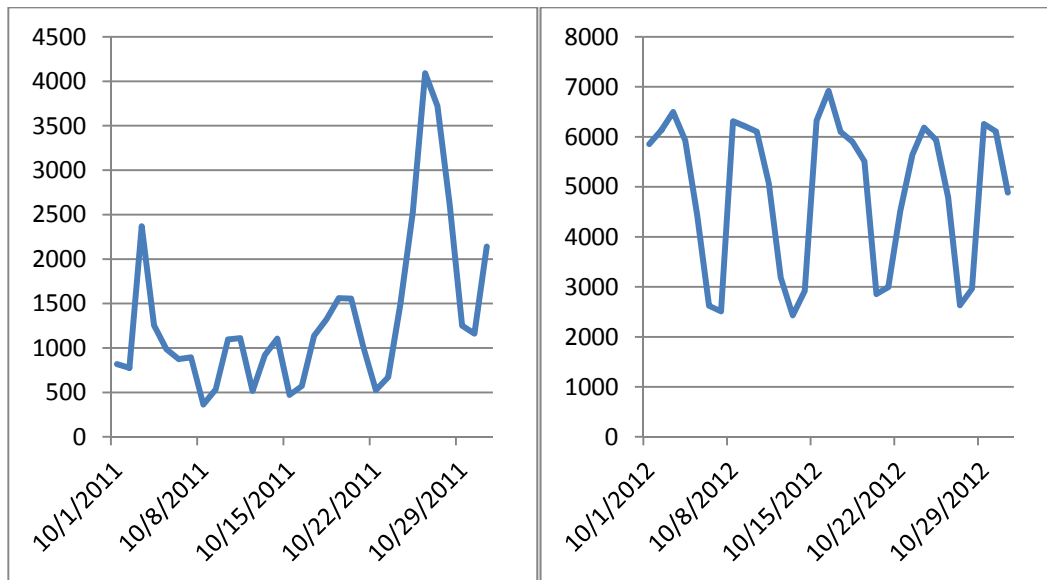


Figura 4 Representación gráfica de los meses de Octubre de 2011 y 2012.

Durante el mes de octubre 2011, el número de consultas realizadas a la base de datos varía entre 500 y 4.000. Este número de peticiones se incrementa notablemente un año después, llegándose a alcanzar en el mismo mes un rango de valores comprendido entre 2.500 y 7.000 accesos. Debido a este aumento en las consultas, se observa una estabilización en el comportamiento de las mismas. Las gráficas pasan de tener una forma de cordillera a una forma de meseta donde se repite un patrón semanal de forma más regular. Además, es evidente y fácilmente observable en las gráficas la diferencia en el número de accesos que existe entre los días laborables y el fin de semana.

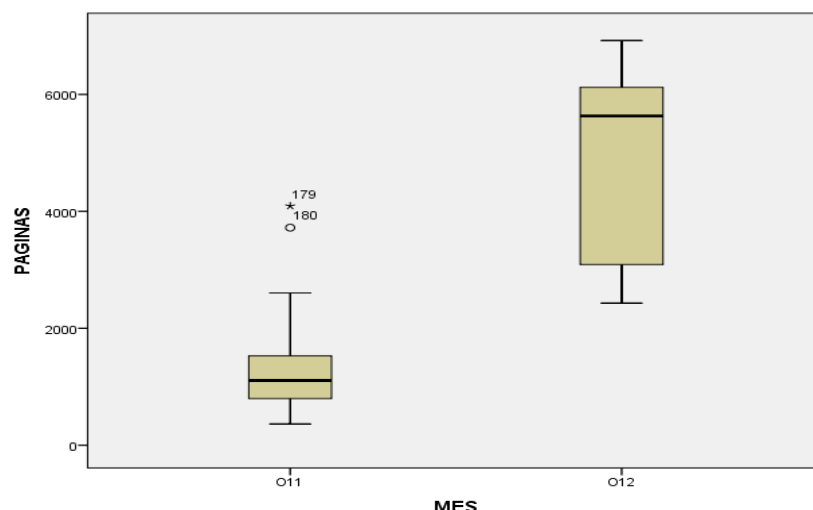


Figura 5 Gráfico de Caja y Bigotes de los meses de Octubre de 2011 y 2012

El diagrama de Caja y Bigotes muestra como los valores mínimos alcanzados en octubre de 2012 comienzan donde terminan los máximos valores alcanzados en el mismo mes del año anterior. Se ve claramente que la dispersión del segundo mes es mucho mayor. Las observaciones 179 y 180 corresponden al día 26-Oct-2011 donde se registraron 4091 consultas y al día 27-Oct-2011 donde se alcanzaron 3271 consultas.

La línea negra dentro de la caja representa la mediana que en el mes de octubre de 2012 aparece claramente desplazada del centro de la caja. Esto obedece a la naturaleza de los datos, ya que durante cinco días laborables se alcanzan valores más altos que durante el sábado y el domingo. La revisión de los datos, previa al análisis detectó la presencia de valores perdidos. En concreto, en esta serie faltan las observaciones correspondientes al 12 y al 13 de noviembre de 2011. Dicha falta se debe a un corte en el suministro de la luz para la realización de actividades de mantenimiento en el cuadro eléctrico del Ministerio. Dichas actuaciones fueron realizadas, como es lógico, durante un fin de semana y corresponden a un sábado y a un domingo.

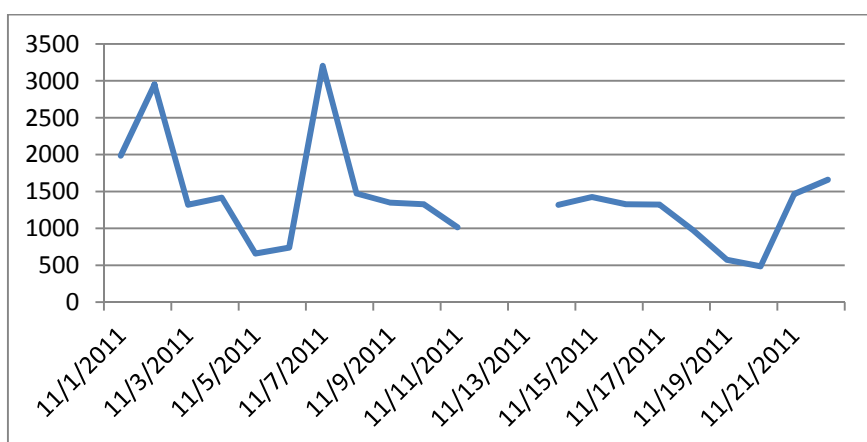


Figura 6 Representación del mes de Noviembre de 2011 con las observaciones perdidas.

Para poder proseguir con el estudio es necesario imputar previamente los valores perdidos de la serie. Existen para ello diversas técnicas y procedimientos que se han aplicado y cuyos valores aparecen en la Tabla. 3 reproducida a continuación.

Día	Páginas	Media Serie	Media Adyacentes	Mediana Adyacentes	Interpol.	Tendencia Lineal
01-nov-11	1986	1986	1986	1986	1986	1986
02-nov-11	2950	2950	2950	2950	2950	2950
03-nov-11	1320	1320	1320	1320	1320	1320
04-nov-11	1417	1417	1417	1417	1417	1417
05-nov-11	658	658	658	658	658	658
06-nov-11	738	738	738	738	738	738
07-nov-11	3204	3204	3204	3204	3204	3204
08-nov-11	1473	1473	1473	1473	1473	1473
09-nov-11	1348	1348	1348	1348	1348	1348
10-nov-11	1328	1328	1328	1328	1328	1328
11-nov-11	1014	1014	1014	1014	1014	1014
12-nov-11		3016	1271	1324	1116	1498
13-nov-11		3016	1271	1324	1217	1509
14-nov-11	1319	1319	1319	1319	1319	1319
15-nov-11	1424	1424	1424	1424	1424	1424
16-nov-11	1328	1328	1328	1328	1328	1328
17-nov-11	1322	1322	1322	1322	1322	1322
18-nov-11	973	973	973	973	973	973
19-nov-11	575	575	575	575	575	575
20-nov-11	486	486	486	486	486	486
21-nov-11	1466	1466	1466	1466	1466	1466

Tabla 3 Datos perdidos junto con sus predicciones.

Los valores perdidos pueden ser sustituidos por la media de la serie que en este caso alcanza el valor de 3.016. Esta solución no es adecuada para series que presentan una marcada tendencia creciente, a menos que estemos imputando en un punto donde la serie crece en torno a su valor medio.

La media y la mediana adyacentes utilizan los valores que preceden y siguen a las observaciones perdidas. En nuestro caso, las observaciones perdidas pertenecen a un fin de semana donde siempre se alcanzan valores significativamente menores que en los días de diario que les preceden o siguen. Por ello, estas dos técnicas de estimación no son adecuadas en este caso.

La interpolación lineal tampoco es adecuada por la misma razón que las anteriores técnicas. Utiliza para su cálculo el último valor válido anterior y el primer valor válido posterior.

Lo mismo ocurre con la tendencia lineal en el punto. De hecho los valores que dan todas estas técnicas para imputar los datos del sábado y el domingo presentan valores propios de días de diario.

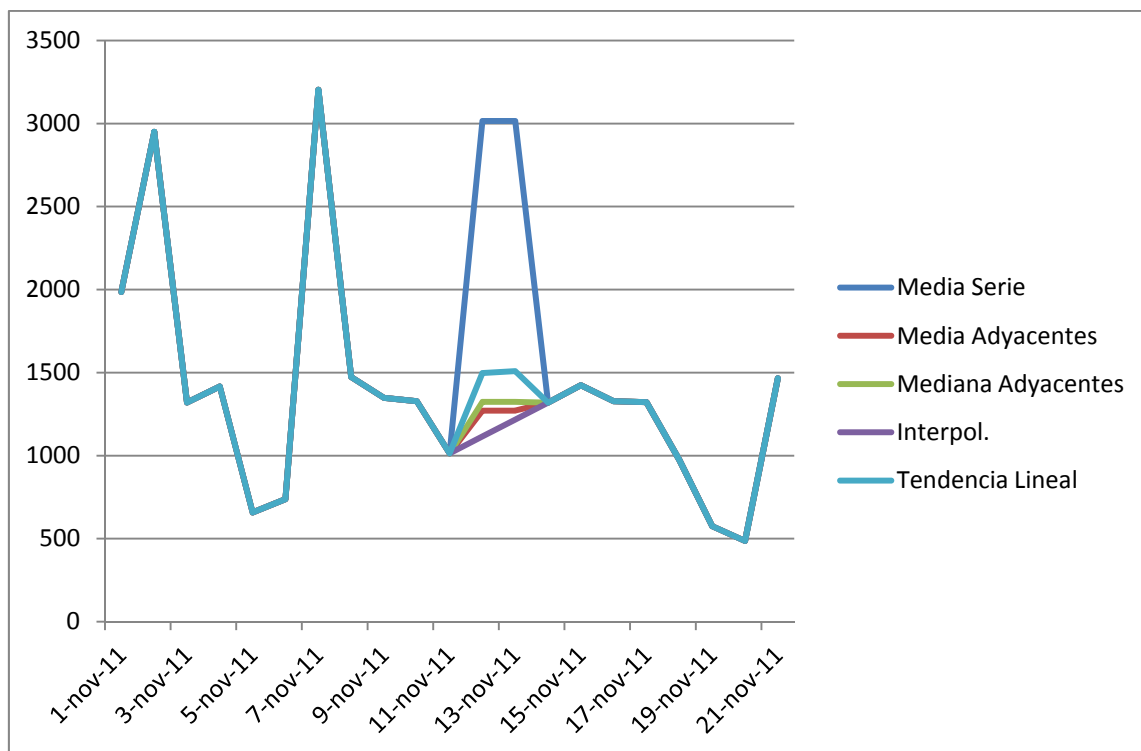


Figura 7 Gráfico con las distintas alternativas para las predicciones.

La gráfica muestra el resultado que se obtendría imputando los valores perdidos por estos cinco métodos. Por el conocimiento de los datos, esperaríamos que la gráfica en este punto presentase un valle y no es así. Se debe de buscar un procedimiento alternativo que nos lleve a una estimación acorde con el comportamiento de la serie. Se toma para ello los valores de las observaciones durante el fin de semana que precede y el siguiente. Con estos datos se calcula la media del sábado y el domingo y se sustituye obteniéndose los valores 617 y 612 recogidos en la siguiente tabla.

Día	Estimación
Sábado 12	617
Domingo 13	612

Tabla 4 Estimación basada en datos del fin de semana.

Sustituyendo estos valores en la gráfica vemos reproducido el comportamiento que esperábamos de la serie, con valores más altos en los días de diario y menores en el fin de semana; tal y como se ha estimado.

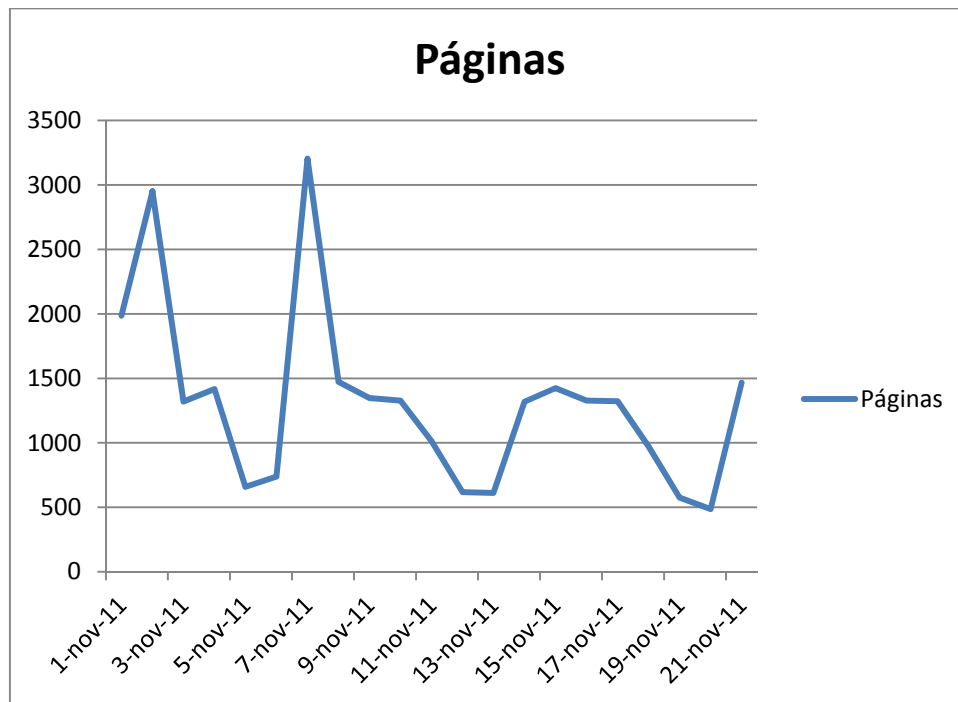


Figura 8 Gráfica del mes de Noviembre de 2011 con las estimaciones de los datos perdidos.

Durante el análisis de la gráfica de la serie se observó la aparición de un pico muy significativo al comienzo de la serie en el mes de marzo de 2011. La figura 9 reproduce a continuación dos gráficas: Una más detallada con el mes en cuestión y otra con los meses que preceden y siguen a dicho mes de marzo.

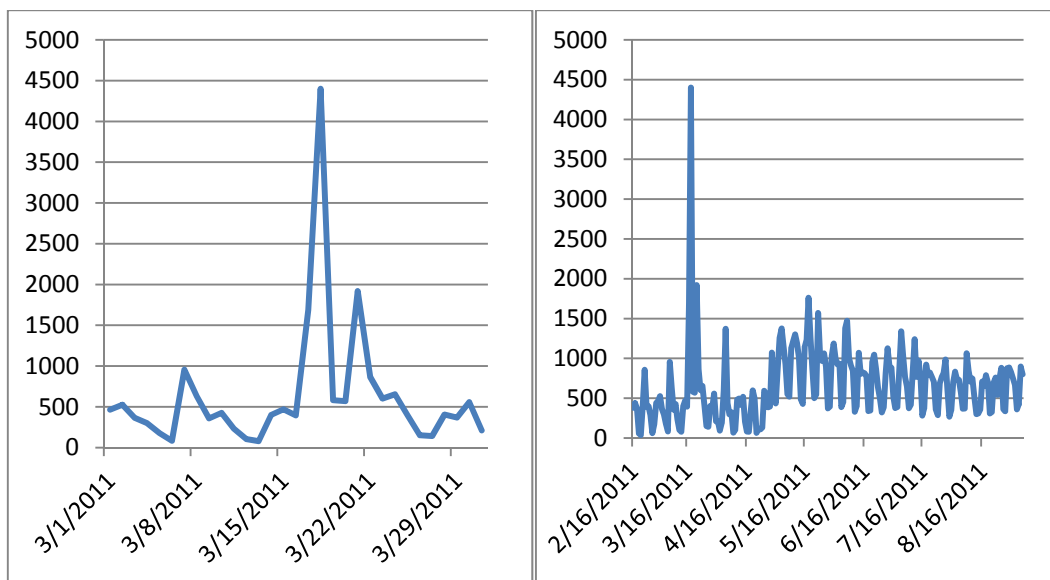


Figura 9 Representación de datos atípicos en el mes de Marzo de 2011, seguida del salto producido en Mayo de 2011.

Hasta esta fecha el número de accesos o consultas a la Base de Datos de Ayudas e Incentivos para Empresas era escaso. Es en este punto cuando comienza la colaboración con los medio de comunicación digitales. Cinco Días incluye en su web las consultas a la base de datos y dedica

un artículo a informar a sus lectores sobre la utilidad y contenidos de la misma así como de su disponibilidad en su web. Esto provoca los picos que aparecen en este mes.

A partir del 1 de Mayo, como se puede ver en la segunda gráfica, de la figura 9, el número de accesos se estabiliza y gracias a dicha colaboración nunca se volverá a bajar de 500 accesos diarios.

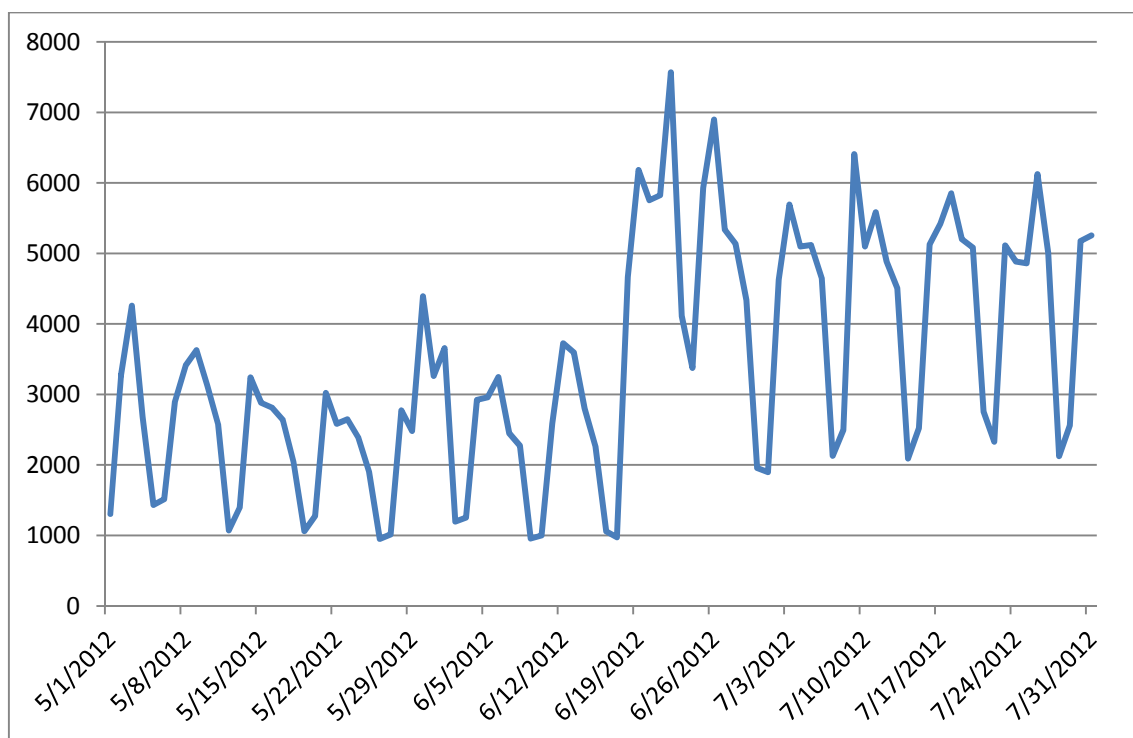


Figura 10. Gráfica del efecto de la incorporación del buscador de consultas de la B.D. de ayudas en el periódico digital Expansión

Otro salto significativo y fácilmente observable en la gráfica se produce en Junio de 2012. Debido esta vez a la aparición de un nuevo colaborador que es la versión digital del diario económico Expansión.

Habíamos observado en la gráfica de la serie que la dispersión de los datos aumentaba con el tiempo. Para homogeneizar la varianza de las observaciones se realiza una transformación logarítmica de los datos. El resultado de dicha transformación puede verse a continuación en la gráfica de la serie ya transformada presente en la figura 11.

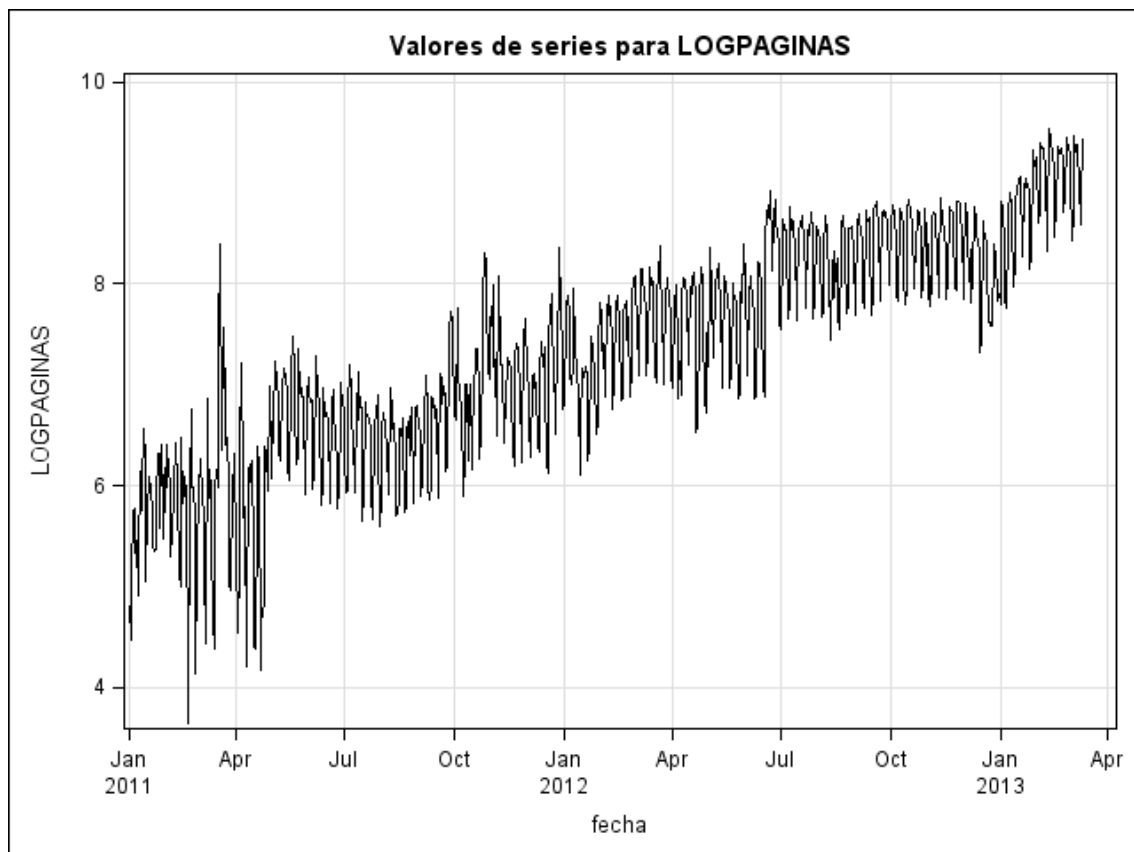


Figura 11 Representación de la serie transformada mediante el logaritmo.

Se puede afirmar que la variabilidad de la serie se estabiliza exceptuando en su comienzo, en concreto los datos anteriores al 1 de Mayo. Esto coincide con lo visto anteriormente. La inclusión de las consultas en el periódico digital Cinco Días hace que el número de accesos aumente significativamente y se estabilice la varianza. Hasta entonces el escaso número de consultas hacía propicio que la variación fuese mucho más acusada.

El período anterior al 1 de Mayo se excluye del estudio por las siguientes razones:

1. Las observaciones iniciales de una serie son las que menos influyen en el modelo.
2. El número de accesos en este período inicial es tan escaso que carece de interés el plantearse cualquier tipo de estudio.
3. En la gráfica se observa dos comportamientos distintos de los datos claramente diferenciados a partir de Mayo de 2011, fecha en la que comienza dicha colaboración con

los medios de información digital.

4. El interés del estudio radica en conocer y modelizar el crecimiento del número de consultas para poder fijar, establecer y proveer suficientes recursos humanos, de hardware y software, capaces de absorber el incremento de la carga de trabajo. Este incremento se produce a raíz de la colaboración con los medios digitales y no antes.

Estudio descriptivo de la serie.

Una vez eliminadas las observaciones del comienzo, la siguiente gráfica muestra la serie con la que vamos a trabajar. Están aquí representados los datos que van desde el 1 de Mayo de 2011 hasta el 11 de marzo de 2013.

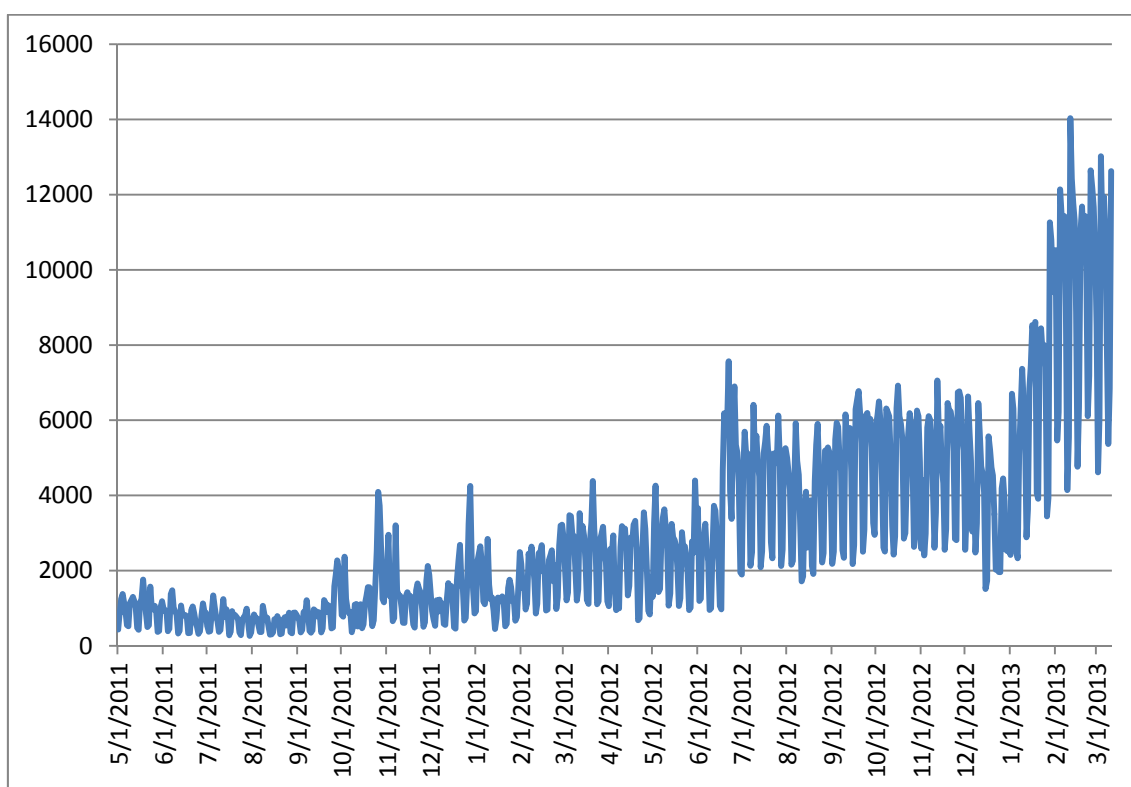


Figura 12 Gráfico de la serie de Consultas a la B.D. de Ayudas desde el 1 de Mayo de 2011.

Ha desaparecido el pico del comienzo pero la serie sigue presentando el mismo aspecto que la serie original, es decir, el gráfico presenta una serie creciente en tendencia y donde aumenta la variabilidad con el tiempo.

Aplicando una diferenciación sobre los datos de la serie obtenemos las gráficas de la función de autocorrelación simple (FAP o ACF) y el gráfico de la función de autocorrelación parcial (FAP o PACF).

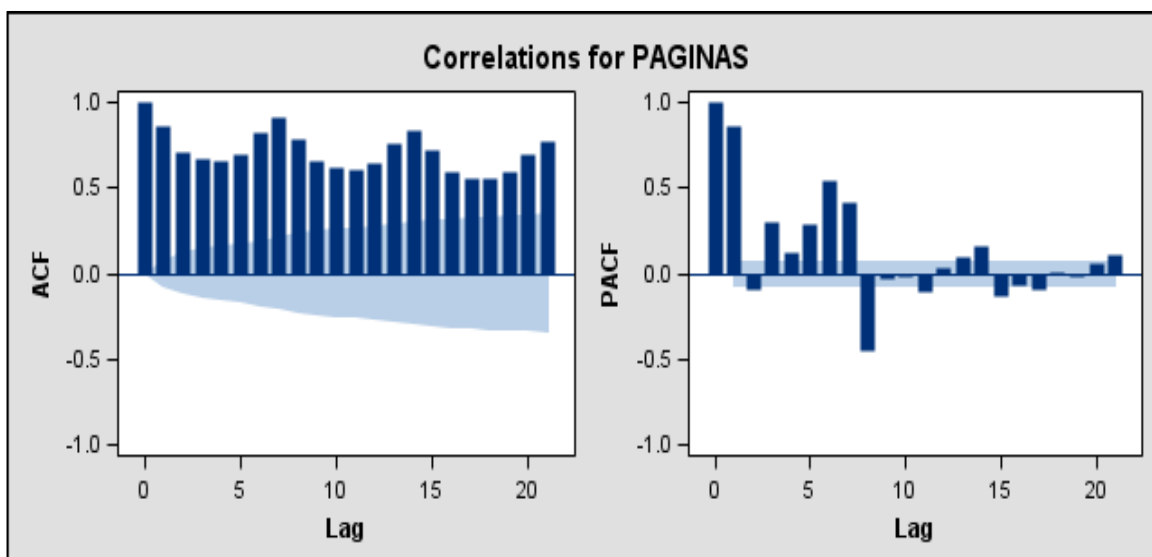


Figura 13 Autocorrelogramas de la serie

El análisis de los gráficos muestra una serie que no es estacionaria. No se observa un decrecimiento rápido o una ruptura de los retardos del FAP o ACF. Sin embargo, se observan picos en el retardo 7 y en sus múltiplos lo que indica que estamos ante una serie estacional que presenta un período de 7 días. Este es un período que cabe esperar de unos datos medidos diariamente y que corresponde a la duración de una semana.

Para determinar con mayor seguridad el período de la serie, disponemos del periodograma. Este gráfico muestra a través de su perfil picos que denotan la existencia de períodos y en su caso de ciclos. La presencia de estos picos identifica los puntos donde se concentra una gran variabilidad de la serie.

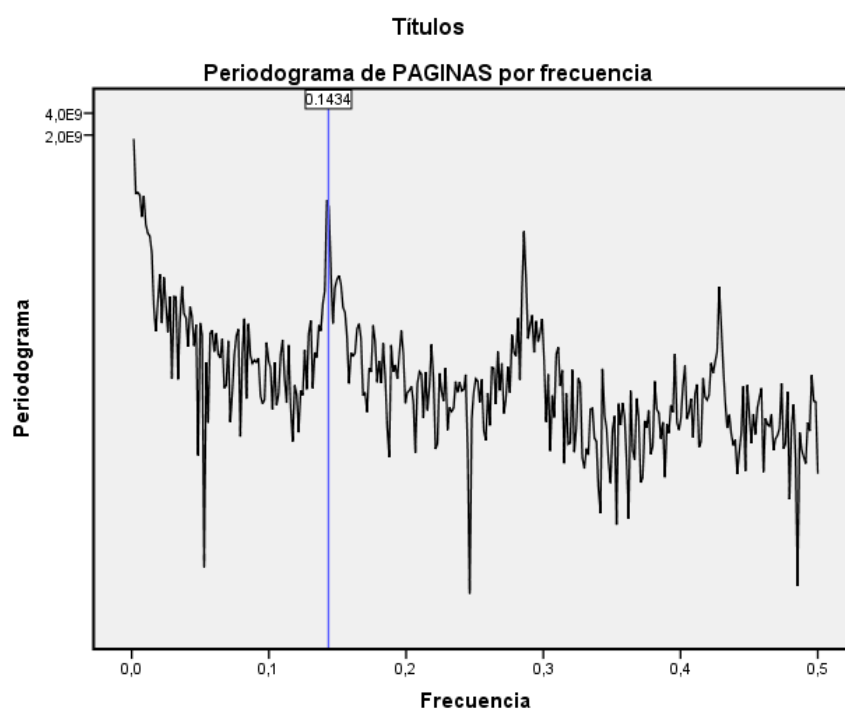


Figura 14 Periodograma de la Serie de las consultas a la B.D.

A la vista de la gráfica presente en la figura 14 se observa un pico en el punto 0,1434. Como el gráfico representa la frecuencia, necesitamos el inverso para obtener el período. Realizando los cálculos obtenemos un período de 7 días o semanal.

$$f = 0,1434 \quad S = \frac{1}{f} = \frac{1}{0,1434} = 6,9635 \approx 7 \text{ días}$$

Para modelar los factores estacionales se opta por el método multiplicativo porque facilita la interpretación de los factores.

En la estacionalidad multiplicativa el componente estacional es un factor por el que se multiplica la serie corregida estacionalmente para dar lugar a la serie original. Los componentes estacionales son proporcionales al nivel general de la serie, de forma que las observaciones sin variación estacional tendrán un componente del 100%. La tabla 4 recoge los factores obtenidos para cada uno de los días de la semana.

Factores estacionales

Período	Factor estacional (%)
Domingo 1	55,1
Lunes 2	123,6
Martes 3	127,3
Miércoles 4	128,7
Jueves 5	116,3
Viernes 6	98,3
Sábado 7	50,7

Tabla 4 Factores estacionales.

El sábado es el día que menor número de accesos se producen. Su proporcionalidad, respecto al número de accesos, se reduce a la mitad. En el extremo opuesto está el miércoles; día en el que el incremento alcanza un valor más alto, llegando a un 28,9%.

El ajuste analítico es un ajuste por regresión de los valores de la serie en función del tiempo. Sirve para hallar la tendencia de la serie mediante la obtención de una función que recoja el comportamiento del fenómeno presente en la serie temporal.

La siguiente tabla aporta la información de los parámetros de las curvas que mejor se ajustan a la serie. Además aporta el coeficiente de determinación R^2 así como el Test de la F de Snedecor.

Resumen del modelo y estimaciones de los parámetros

Variable dependiente: Ciclo de tendencias para PAGINAS

Ecuación	Resumen del modelo					Estimaciones de los parámetros			
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1	b2	b3
Lineal	,760	2157,903	1	680	,000	-609,251	10,668		
Cuadrático	,848	1891,383	2	679	,000	990,336	-3,363	,021	
Cúbico	,855	1334,298	3	678	,000	439,619	6,278	-,015	3,442E-005
Compuesto	,892	5591,559	1	680	,000	584,008	1,004		
Crecimiento	,892	5591,559	1	680	,000	6,370	,004		
Exponencial	,892	5591,559	1	680	,000	584,008	,004		
Logística	,892	5591,559	1	680	,000	,002	,996		

Tabla 5 Resumen de las funciones que mejor se adaptan al número de consultas accesos a la B.D.

El valor máximo obtenido para el coeficiente R^2 es igual a 0,892. Alcanzan este valor cuatro de los modelos; que son Compuesto, Crecimiento, Exponencial y Logística. Al ser igual de buenos, elegimos uno de ellos, en concreto el Modelo Compuesto.

$$Z(t) = a * b^t$$

$$Z(t) = 584,008 * 1,004^t$$

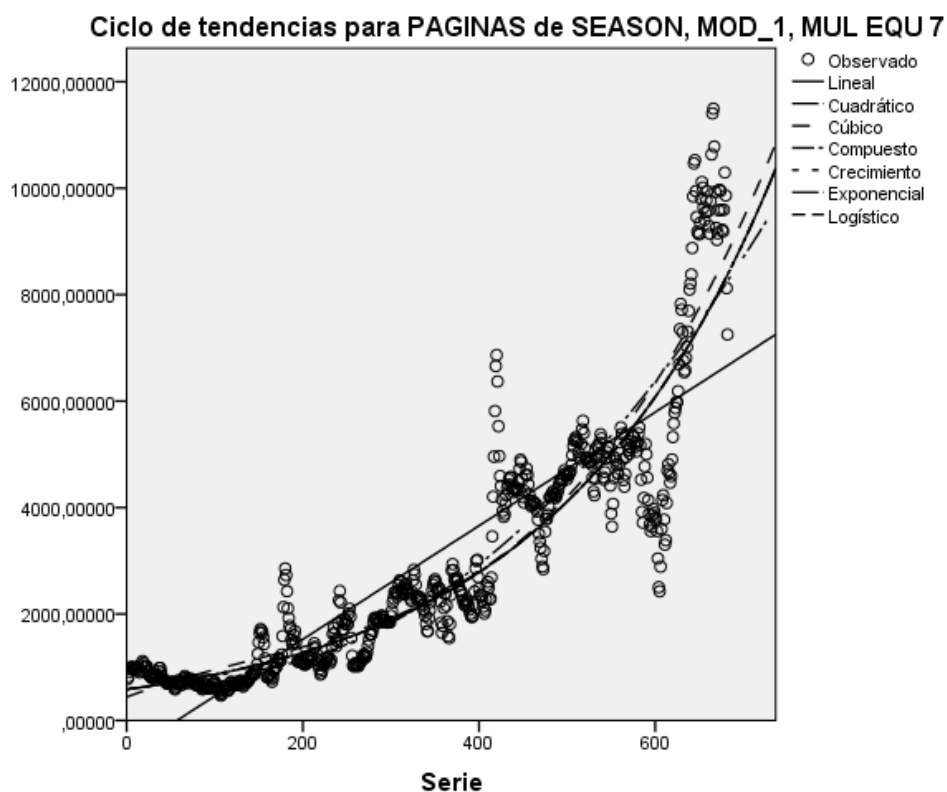


Figura 15 Ajuste de las curvas a la nube de puntos.

Ajuste y predicciones.

Para estabilizar la dispersión creciente en el tiempo y homogeneizar la varianza se procede a realizar una transformación logarítmica de los datos originales. El resultado de dicha transformación puede observarse en la siguiente gráfica.

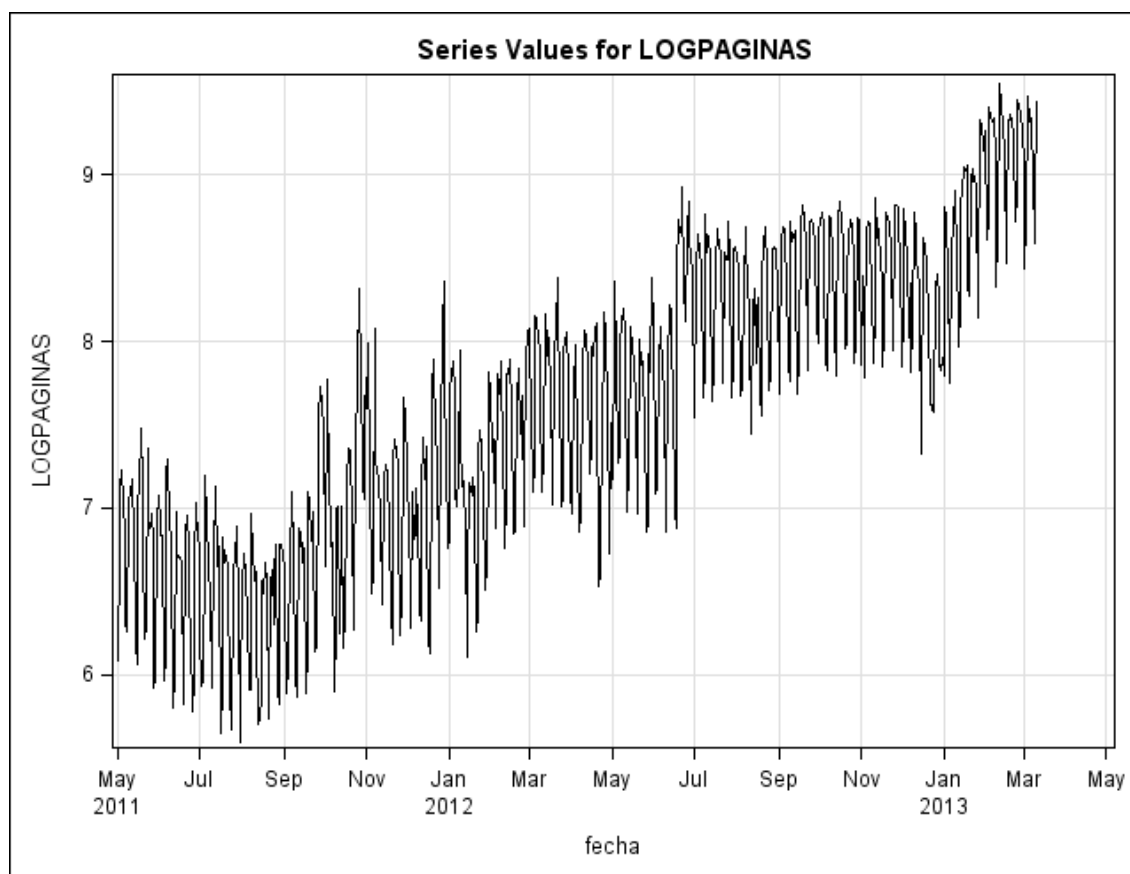


Figura 16 Representacion del gráfico de la serie transformada mediante el logaritmo.

Al contrario que en la serie original la dispersión no aumenta ni disminuye con el tiempo, por lo que se da por buena la transformación logarítmica de los datos como paso previo a la obtención de un modelo ARIMA.

Debido a la presencia de la tendencia creciente de los datos se hace necesario aplicar una diferenciación de los mismos con objeto de estabilizar dicha tendencia. La figura 17 muestra los gráficos de las funciones de autocorrelación simple y parcial.

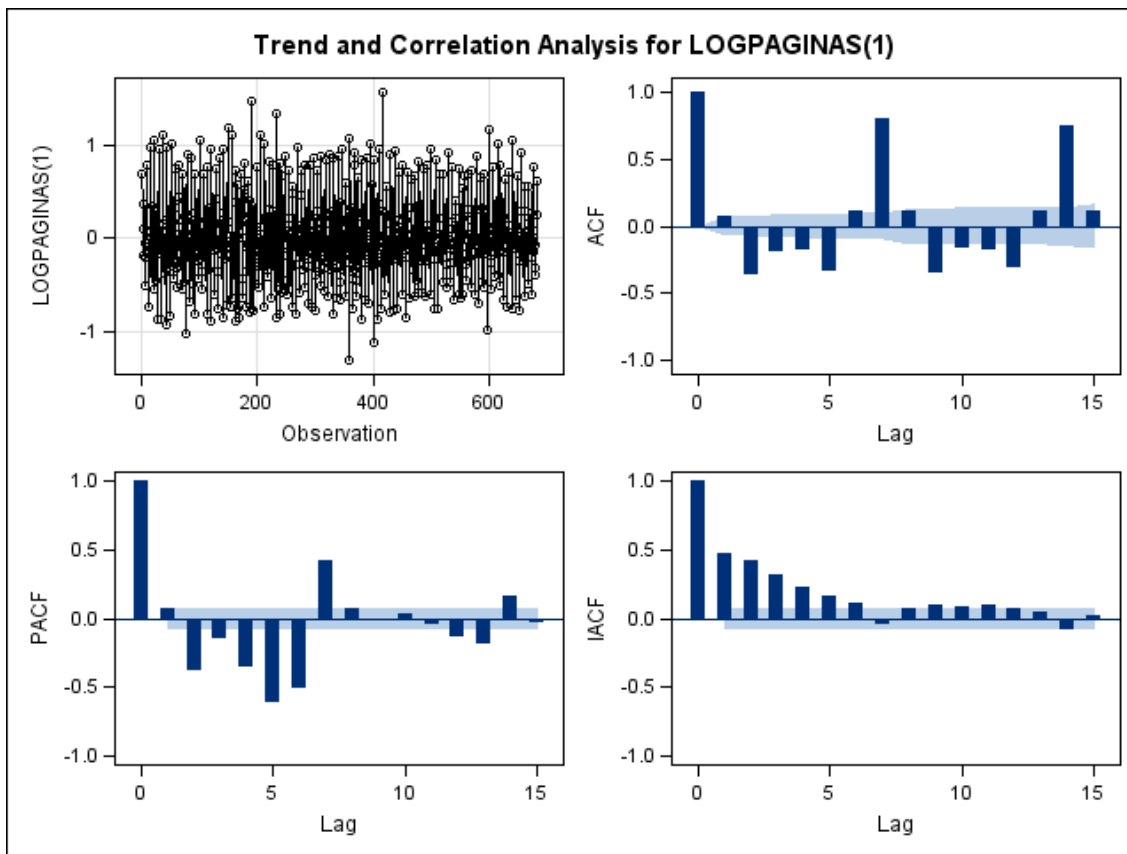


Figura 17 Autocorelogramas del Modelo ARIMA (0,1,0) aplicados a la serie transformada por el logaritmo.

Se sabía por el análisis del periodograma que los datos presentaban un período semanal. Los correlogramas simple y parcial presentan picos en el retardo 7º y en sus múltiplos.

Aplicamos junto a la anterior diferenciación una diferenciación estacional con período de 7.

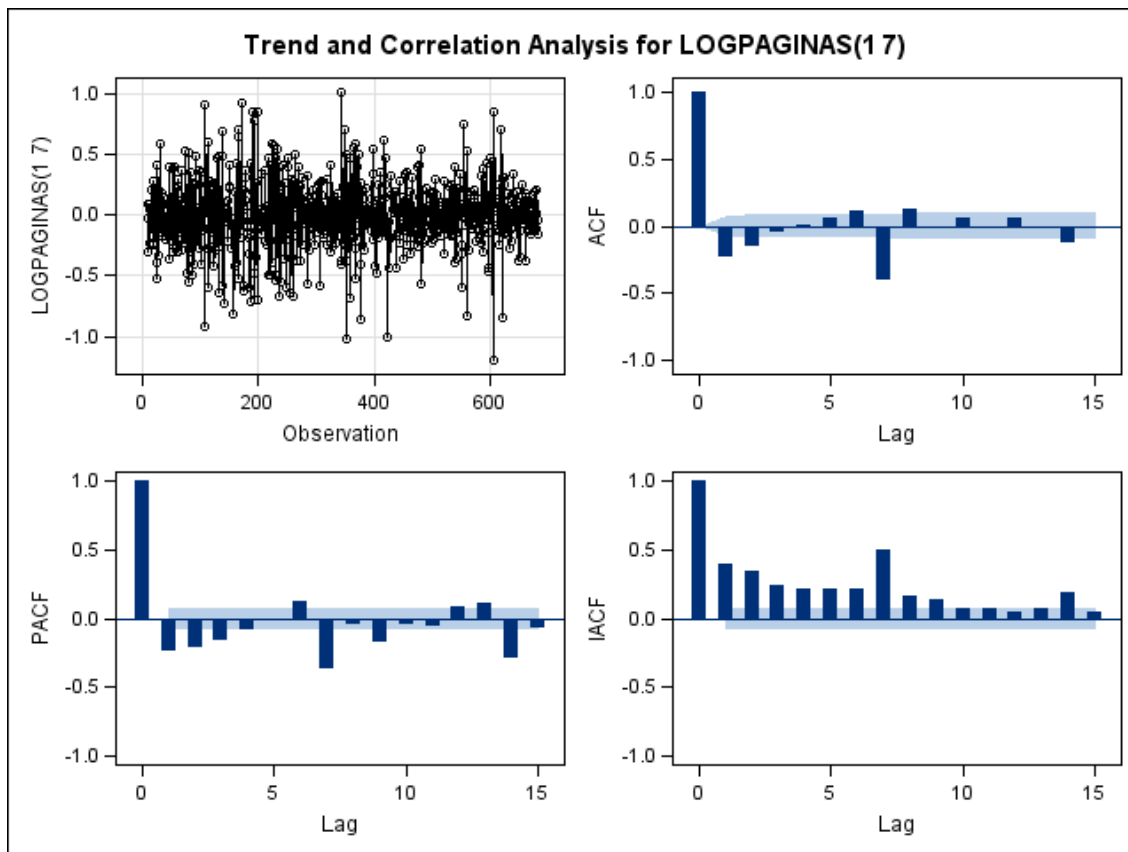


Figura 18 Autocorelogramas del Modelo ARIMA (0,1,0)(0,1,0)₇ aplicados a la serie transformada por el logaritmo.

Los correlogramas siguen mostrando un coeficiente significativo en el retardo de 7.

Aparte de lo anterior aplicamos ahora el orden 7 en la parte autoregresiva estacional cuyo resultado se muestra más adelante.

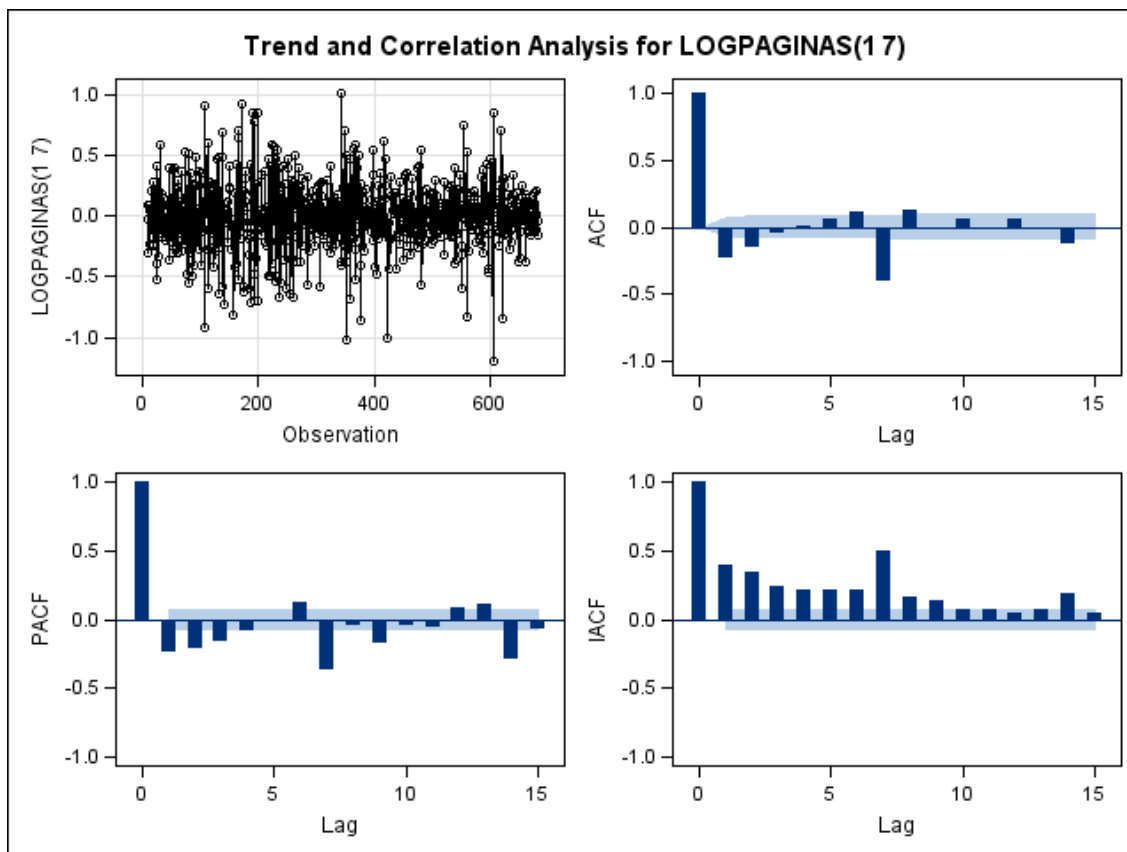


Figura 19 Autocorelogramas del Modelo ARIMA (0,1,0)(1,1,0)₇ aplicados a la serie transformada por el logaritmo.

Se siguen observando dentro de los diagramas de las autocorrelaciones un pico significativo en el retardo séptimo y en su múltiplo catorce.

Añadimos al modelo anterior el orden 7 en la parte de medias móviles de la parte estacional mostrándose a continuación las gráficas obtenidas.

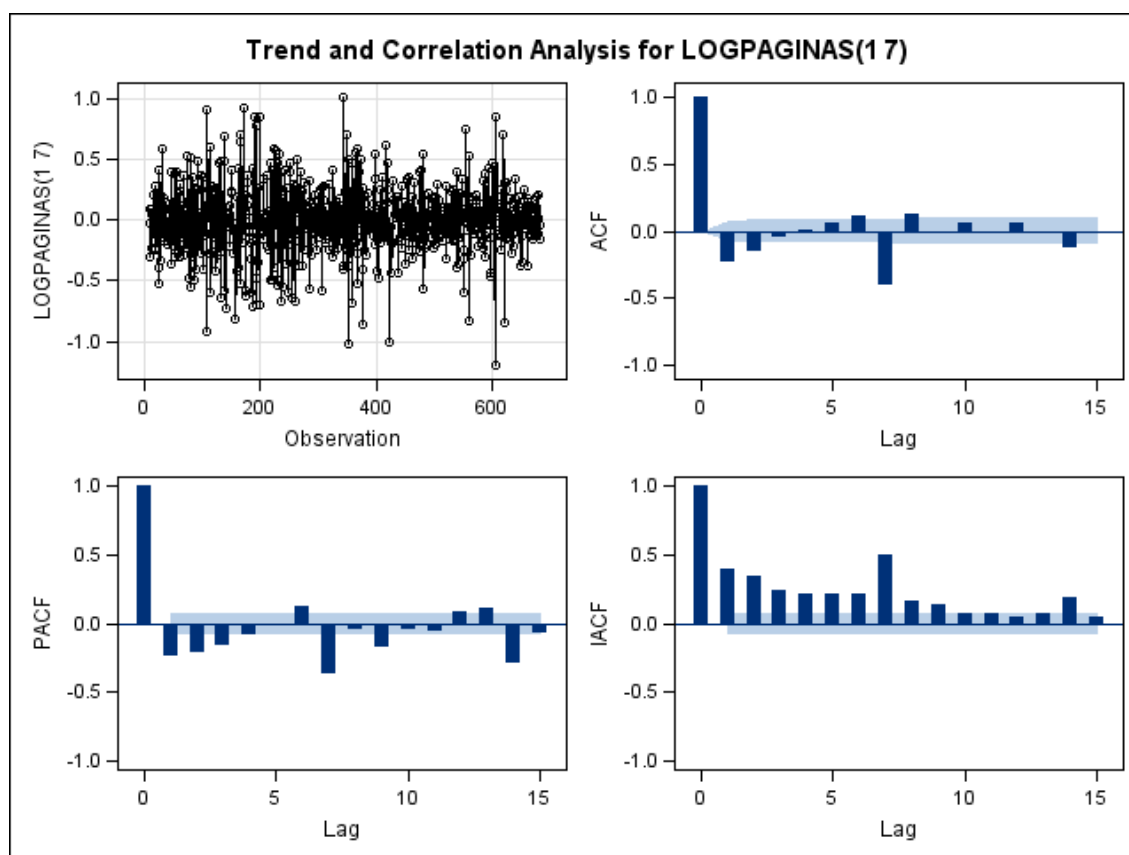


Figura 20 Autocorelogramas del Modelo ARIMA (0,1,0)(1,1,1)₇ aplicados a la serie transformada por el logaritmo.

Igual que con los anteriores modelos el retardo séptimo se distingue en el perfil de las gráficas.

Uno de los problemas frecuentes en el ajuste de un modelo ARIMA es la sobreparametrización. Conviene no solamente añadir parámetros, sino también considerar la posibilidad de sacar del modelo o eliminar alguno de ellos. Eliminamos la diferenciación del orden 1 y volvemos a calcular el modelo para la obtención de las gráficas de los autocorrelogramas de la serie.

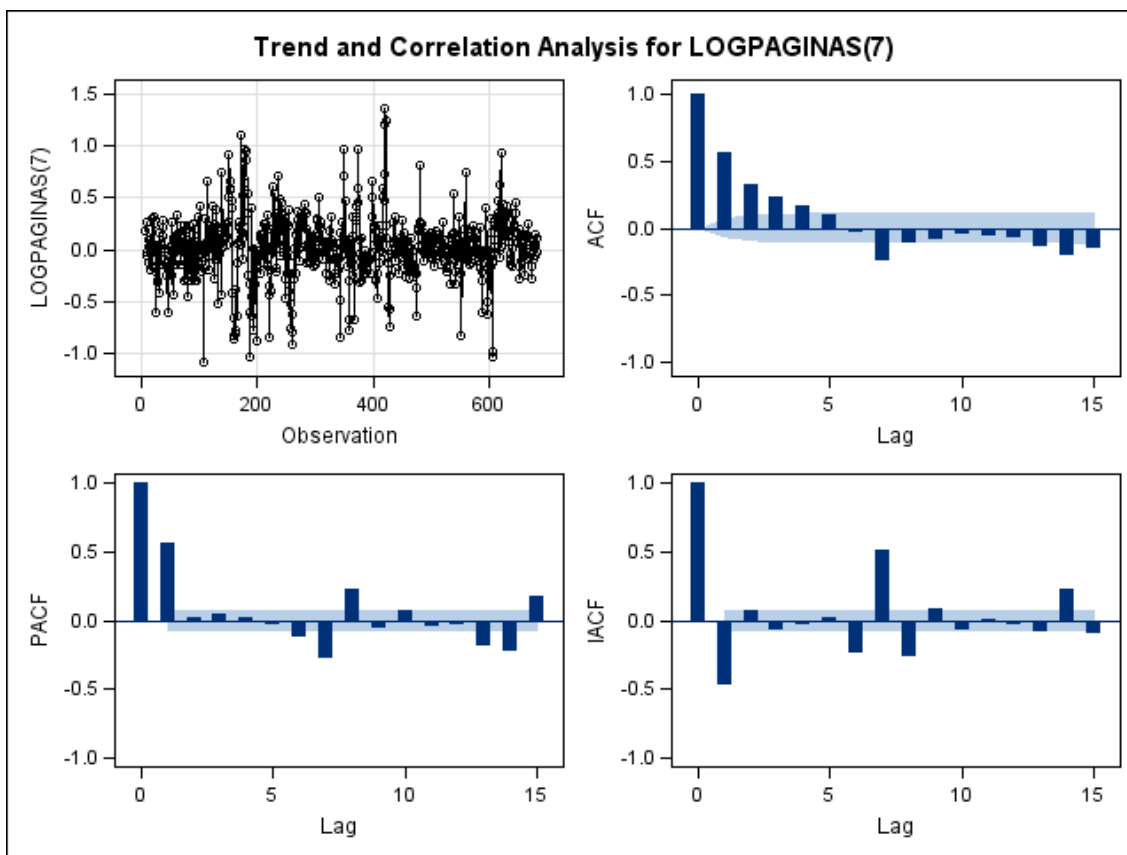


Figura 21 Autocorelogramas del Modelo ARIMA (0,0,0)(1,1,1)₇ aplicados a la serie transformada por el logaritmo.

Observamos ahora en el gráfico de autocorrelación simple un decrecimiento rápido de los coeficientes, propio de las series que son estacionarias. Sin embargo, el retardo séptimo a pesar de ser más pequeño que en casos anteriores sigue destacando.

Incluimos en el modelo anterior el orden 1 autoregresivo. Se vuelve a calcular el modelo y se obtienen los siguientes resultados.

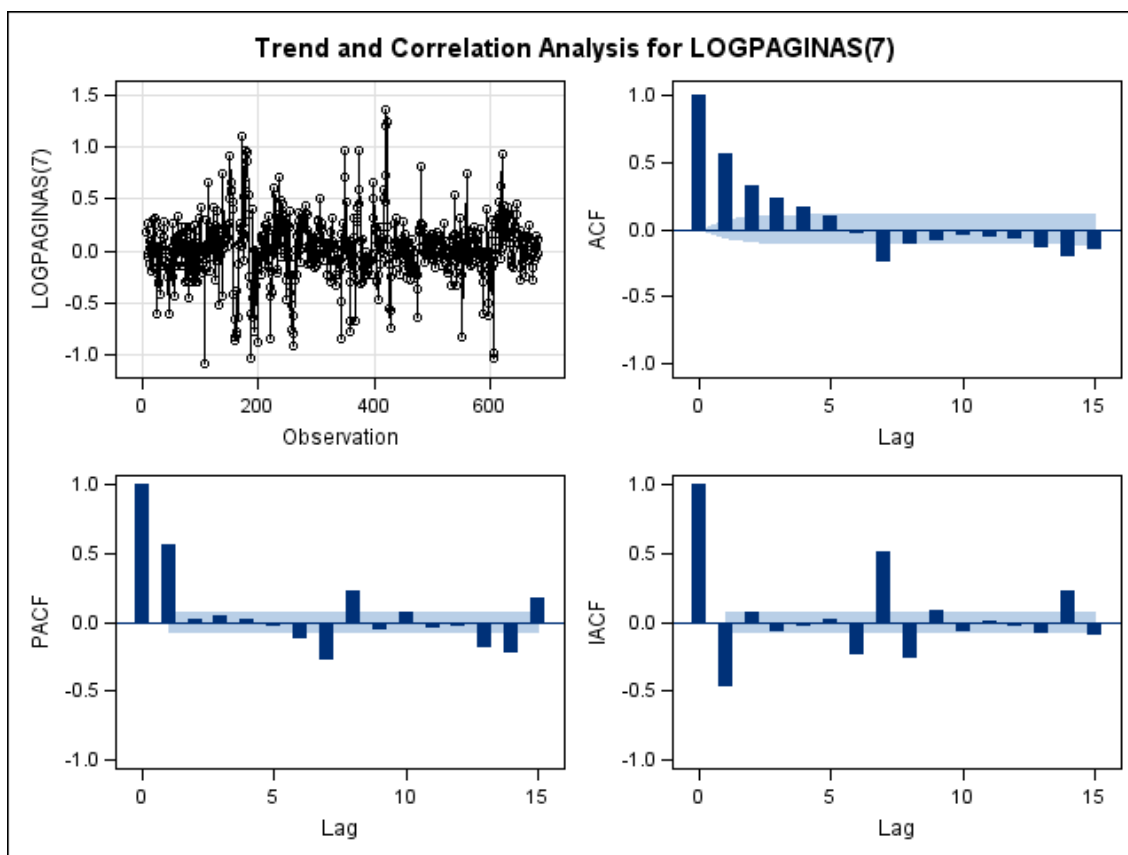


Figura 22 Autocorelogramas del Modelo ARIMA (1,1,0)(1,1,1)₇ aplicados a la serie transformada por el logaritmo.

Los gráficos de autocorrelaciones de los residuos muestran los coeficientes de autocorrelación dentro de los límites del intervalo de confianza. Por ello se acepta el modelo ARIMA (1,0,0)(1,1,1)₇ como idóneo.

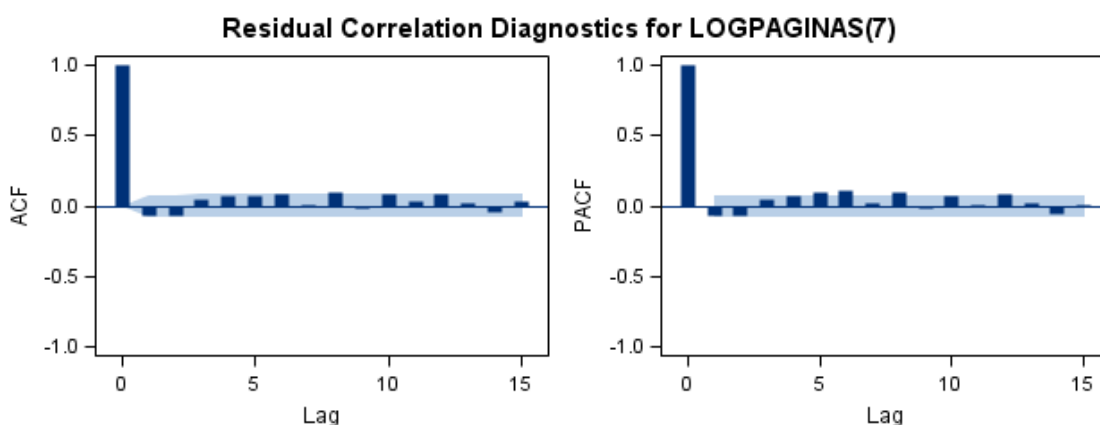


Figura 23 Autocorelogramas de los residuos para el Modelo ARIMA (0,1,0)(0,1,0)₇ aplicados a la serie transformada por el logaritmo.

Una vez obtenido el modelo hay que detectar si existen singularidades o datos atípicos dentro de la serie. Los investigadores saben que en junio de 2012 se facilitó al diario digital Expansión la inclusión del buscador de la Base de Datos de Ayudas e Incentivos de las Empresas. En la gráfica de los datos se puede observar que en esa fecha aparece un escalón incrementándose el número de accesos registrados. Hay que determinar ahora si ese escalón es o no significativo dentro de nuestro modelo.

Se solicita al software la detección de saltos y se obtiene la siguiente tabla.

Outlier Details					
Obs	Time ID	Type	Estimate	Chi-Square	Approx Prob> ChiSq
24	24-MAY-2011	Shift	-0.49254	20.16	<.0001
613	02-JAN-2013	Shift	0.29007	12.87	0.0003
551	01-NOV-2012	Shift	-0.27966	16.93	<.0001
415	18-JUN-2012	Shift	0.25568	18.70	<.0001
604	24-DEC-2012	Shift	-0.24146	9.56	0.0020

Tabla 6 Relación de posibles saltos en la serie.

Como se esperaba por el conocimiento previo de las observaciones aparece un salto en la fecha del 18 de junio de 2012, coincidiendo con la entrada del periódico digital.

Los saltos de diciembre de 2012 seguido del de enero de 2013 no se consideran como saltos en sí. Son debidos al descenso de la actividad por las vacaciones navideñas. El investigador sospechará la presencia de un posible ciclo anual. Para poder estudiar el comportamiento anual de la serie, se requiere de un número suficiente de años. En este momento solo se dispone de observaciones correspondientes a dos años, por lo que no es posible establecer ciclos anuales.

Observando la gráfica de la serie se puede descartar los saltos de mayo de 2011 y de noviembre de 2012.

Incluimos en el modelo el salto correspondiente al 18 de junio de 2012 obteniéndose las siguientes tablas que muestran la estimación de los parámetros por el método de Máxima Verosimilitud, así como la correlación entre los parámetros estimados del modelo.

Correlations of Parameter Estimates						
Variable Parameter	LOGPAGINAS MU	LOGPAGINAS MA1,1	LOGPAGINAS AR1,1	LOGPAGINAS AR2,1	SHJunio2012 NUM1	
LOGPAGINAS MU	1.000	-0.003	0.011	-0.011	-0.795	
LOGPAGINAS MA1,1	-0.003	1.000	-0.066	0.311	0.019	
LOGPAGINAS AR1,1	0.011	-0.066	1.000	-0.246	-0.013	
LOGPAGINAS AR2,1	-0.011	0.311	-0.246	1.000	0.016	
SHJunio2012 NUM1	-0.795	0.019	-0.013	0.016	1.000	

Tabla 7 Correlaciones de los parámetros estimados del modelo incluyendo la variable salto SHJunio2012.

Se observa que el parámetro que incluye el salto en el modelo presenta una correlación muy alta (-0,795) con el término independiente, lo que indica que una gran parte de lo que aporta este parámetro ya está contenida en dicho término.

Maximum Likelihood Estimation						
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag	Variable
MU	0.02762	0.0023000	12.01	<.0001	0	LOGPAGINAS
MA1,1	0.97880	0.01952	50.15	<.0001	7	LOGPAGINAS
AR1,1	0.71957	0.02735	26.31	<.0001	1	LOGPAGINAS
AR2,1	0.18725	0.04102	4.56	<.0001	7	LOGPAGINAS
NUM1	-0.0010481	0.0051904	-0.20	0.8400	0	SHJunio2012

Tabla 8 Parámetros del modelo estimado. Se incluye la variable salto SHJunio2012.

En concordancia con la tabla de las correlaciones de los parámetros, el test de la t de Student rechaza la inclusión de la variable salto dándole un p-valor de 0,84. Por ello, no incluiremos dicho salto en nuestro modelo.

Una comparativa de la bondad de ajuste de los distintos modelos calculados está presente en la tabla 9 que recoge la estimación de la varianza, la desviación típica estimada, el criterio de información de Akaike y por último el criterio bayesiano de Schwarz.

S = 7	ARIMA (0,1,0)(0,1,0)	ARIMA (0,1,0)(1,1,0)	ARIMA (0,1,0)(1,1,1)	ARIMA (0,0,0)(1,1,1)	ARIMA (1,0,0)(1,1,1)
Variance Estimate	0.083651	0.07025	0.082962	0.07892	0.041813
Std Error Estimate	0.289225	0.265046	0.288032	0.280927	0.202703
AIC	241.1097	124.6021	236.8903	204.2195	-214.738
SBC	245.6214	133.6255	245.9167	217.7591	-196.685
Nº of Residuals	673	673	674	675	674

Tabla 9 Comparación de las medidas de la bondad de ajuste de los distintos modelos calculados.

La tabla recoge como hemos ido ajustando el modelo. La varianza y la desviación típica han ido reduciéndose. Asimismo, los criterios de Akaike y de Schwarz han ido alcanzando paulatinamente valores cada vez más pequeños, lo que demuestra que el último de los modelos obtenidos es mejor que cualquiera de los que le preceden.

Quedaría ahora determinar si los residuos se ajustan o no a una distribución normal que es lo que cabría esperar en un ajuste perfecto de datos.

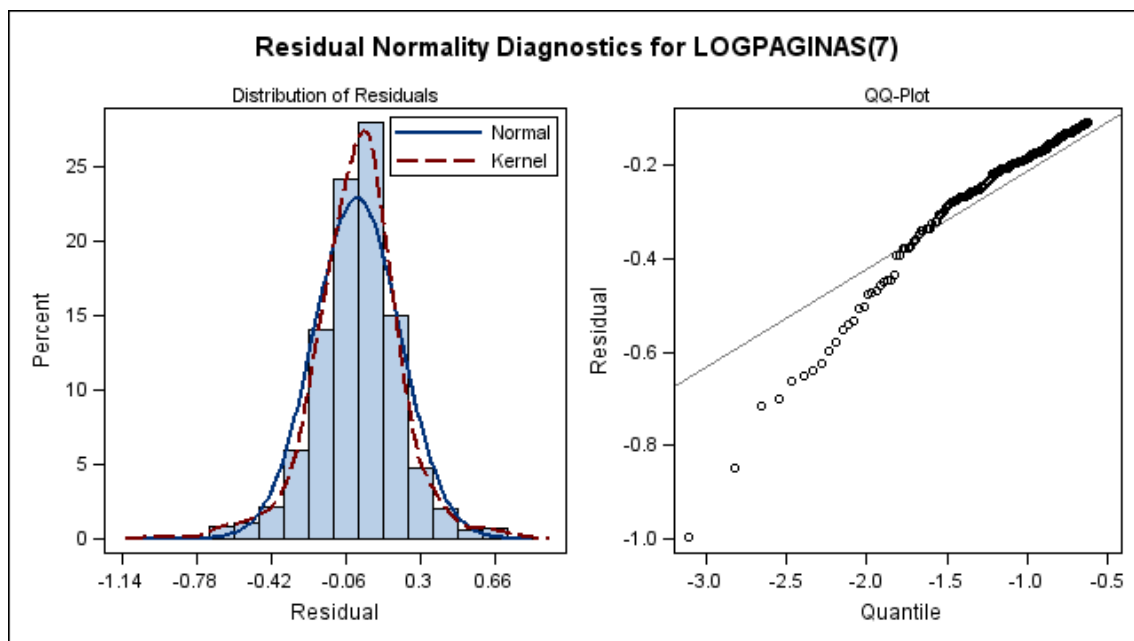


Figura 24 Gráficos de normalidad de los residuos.

El histograma muestra que la distribución de los residuos es algo más apuntada que la distribución de la normal. El gráfico QQ-Plot no muestra un ajuste perfecto a la recta; ajuste que ocurriría si los residuos siguiesen una distribución normal.

Es cierto que hemos concluido que los datos tienen una periodicidad de 7 días, que los gráficos de autocorrelación de los residuos están dentro de los límites de confianza y que nuestro modelo es el que mejor se ajusta de todos los calculados. Sin embargo, no estamos ante un modelo de libro o de manual, donde todo ajusta a la perfección. Nos encontramos ante un modelo real en el que la serie sigue el comportamiento del modelo obtenido aunque no todos los resultados obtenidos sean ideales.

El valor de los parámetros estimado para nuestro modelo según el método de Máxima Verosimilitud está recogido en la tabla. 10.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	0.02725	0.0014033	19.42	<.0001	0
MA1,1	0.97847	0.01927	50.79	<.0001	7
AR1,1	0.71957	0.02733	26.33	<.0001	1
AR2,1	0.18725	0.04099	4.57	<.0001	7

Tabla 10 Estimación de los parámetros del modelo.

El valor del estadístico de la t de Student y su p-valor asociado lleva a rechazar para todos los parámetros la hipótesis nula y por tanto se aceptan todos como significativos dentro del modelo.

Los valores de las correlaciones entre los distintos parámetros se muestran en la siguiente tabla.

Correlations of Parameter Estimates				
Parameter	MU	MA1,1	AR1,1	AR2,1
MU	1.000	0.044	-0.001	0.010
MA1,1	0.044	1.000	-0.064	0.311
AR1,1	-0.001	-0.064	1.000	-0.246
AR2,1	0.010	0.311	-0.246	1.000

Tabla 11 Correlación de los parámetros del modelo.

Todos los valores de las correlaciones son en general pequeños, lo que es deseable, a excepción del parámetro autoregresivo estacional que presenta unos valores de 0,3 con el de medias móviles estacional y de 0,24 con el autoregresivo de orden 1. No son valores muy altos pero tampoco son tan pequeños como se desearía. Pero como se ha visto anteriormente nos encontramos ante un modelo real y no ideal.

Estimated Mean	0.027249
Period(s) of Differencing	7

Autoregressive Factors

Factor 1:	1 - 0.71957 B**(1)
Factor 2:	1 - 0.18725 B**(7)

Moving Average Factors

Factor 1:	1 - 0.97847 B**(7)
-----------	--------------------

Tabla 12 Modelo ARIMA (1,0,0)(1,1,1)₇

El modelo obtenido seria el siguiente:

$$(1 - 0,187 B^7) (1 - 0,719 B) (1 - B^7) Y_t = 0,027 + (1 - 0,978 B^7) E_t$$

siendo $Y_t = \log X_t$

La figura 25 muestra la serie del logaritmo de los datos a la que se superponen los valores predichos y los límites de confianza. Al final de la misma se añaden las predicciones para el mes siguiente.

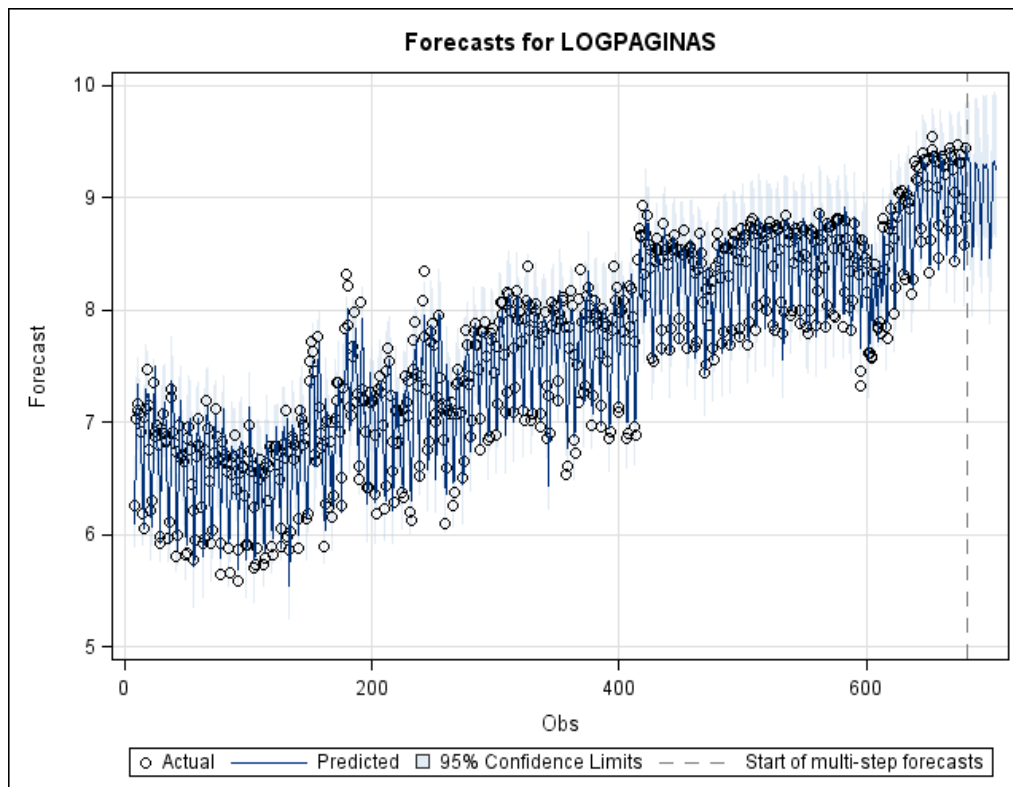


Figura 25 Gráfico de la serie transformada mediante logaritmo representada con los intervalos de confianza y las predicciones.

Como puede observarse en la gráfica, existe una superposición entre las observaciones y las estimaciones. Interesa ahora conocer el aspecto que presentan las predicciones para el mes siguiente. A continuación se muestra la gráfica de las predicciones para ese intervalo de tiempo.

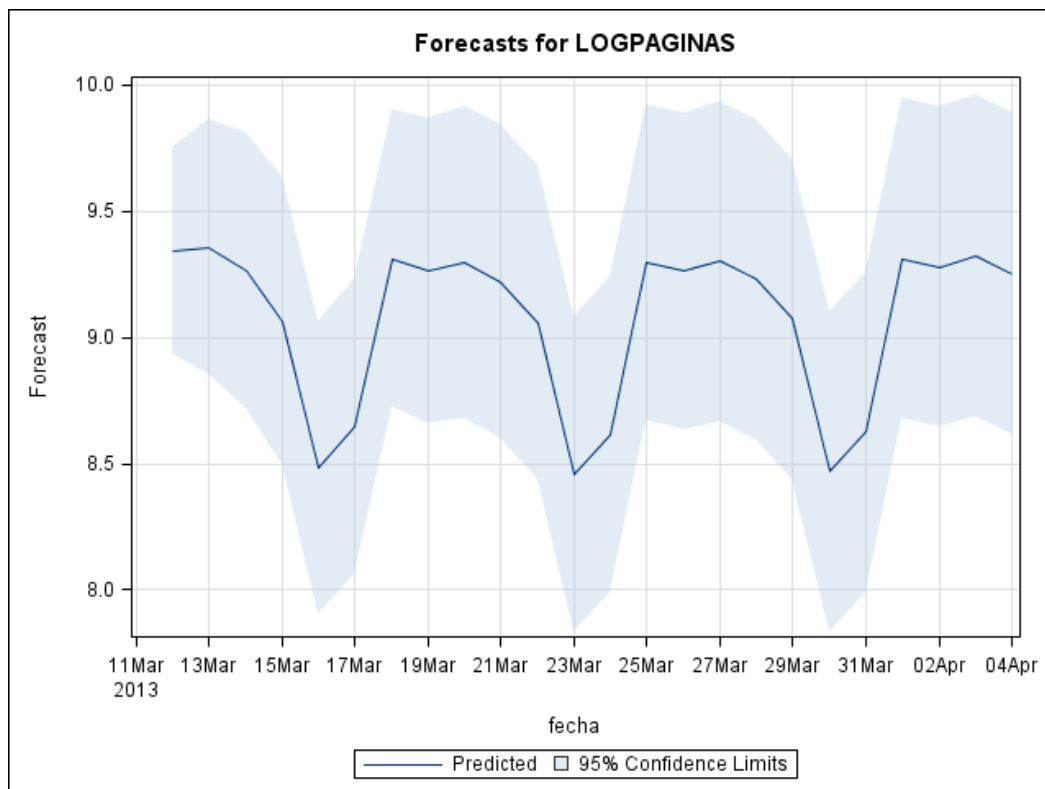


Figura 26 Detalle de las predicciones

Vemos el comportamiento que esperábamos. Las estimaciones correspondientes a los días laborables son significativamente más altas que las pertenecientes al fin de semana. La gráfica obtenida recuerda mucho a la gráfica del mes de octubre de 2012 que analizábamos al comienzo de este trabajo.

El siguiente gráfico representa la serie original sin la transformación logarítmica. A ella se superponen los intervalos de confianza del 95% y se añaden las predicciones del mes siguiente representadas al final.

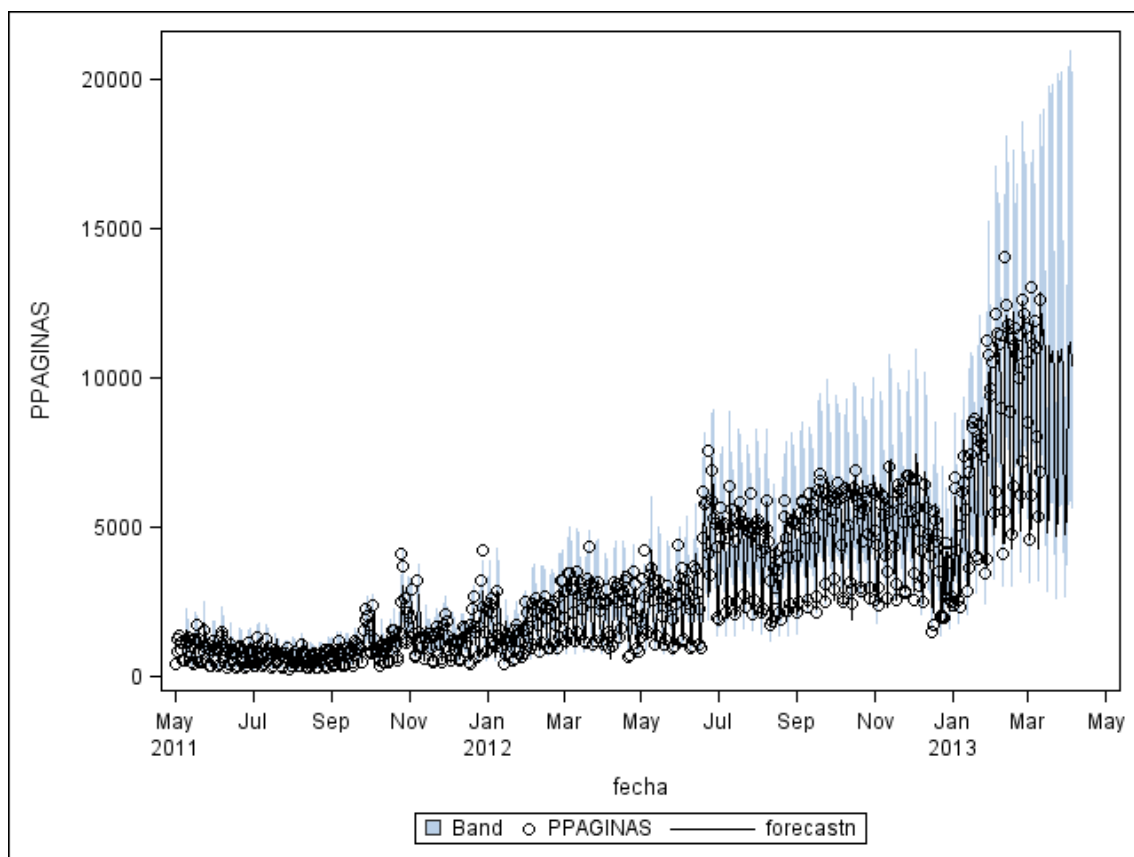


Figura 27 Serie de consultas a la B.D. de Ayudas con los intervalos de confianza y las predicciones.

Volvemos a ver nuestra serie original con la tendencia y variabilidad creciente y en consecuencia ocurre lo mismo con los intervalos de confianza.

Por último, la tabla 13 recoge las predicciones de las cuatro semanas siguientes a la finalización de la serie. Estas predicciones están acompañadas de los límites del intervalo de confianza al 95%.

Resumen de predicciones para el mes siguiente.

Obs	fecha	PPAGINAS	l95n	u95n	forecastn
677	07-mar-13	11031	7038.71	15989.45	10608.73
678	08-mar-13	8031	6064.38	13776.12	9140.22
679	09-mar-13	5371	2862.46	6502.50	4314.30
680	10-mar-13	6841	3991.63	9067.55	6016.17
681	11-mar-13	12621	8528.78	19374.35	12854.56
682	12-mar-13		7588.99	17239.48	11438.11
683	13-mar-13		6995.50	19214.31	11593.69
684	14-mar-13		6095.24	18231.38	10541.57
685	15-mar-13		4898.11	15271.81	8648.87
686	16-mar-13		2708.16	8621.74	4832.08
687	17-mar-13		3170.28	10200.78	5686.76
688	18-mar-13		6157.39	19920.47	11075.11
689	19-mar-13		5763.20	19367.85	10565.07
690	20-mar-13		5899.47	20209.67	10919.08
691	21-mar-13		5437.15	18809.12	10112.76
692	22-mar-13		4612.67	16037.31	8600.86
693	23-mar-13		2522.06	8791.40	4708.76
694	24-mar-13		2948.83	10292.75	5509.22
695	25-mar-13		5827.96	20356.21	10891.98
696	26-mar-13		5608.33	19755.27	10525.88
697	27-mar-13		5820.19	20590.71	10947.23
698	28-mar-13		5412.96	19192.85	10192.65
699	29-mar-13		4636.83	16459.89	8736.23
700	30-mar-13		2531.62	8992.11	4771.22
701	31-mar-13		2960.81	10519.80	5580.96
702	01-apr-13		5871.77	20865.81	11068.84
703	02-apr-13		5675.41	20268.06	10725.19
704	03-apr-14		5903.24	21135.43	11169.93
705	04-apr-14		5498.91	19713.69	10411.72

Tabla 13 Resumen de las predicciones

Conclusiones y análisis futuros.

Una vez obtenido el modelo que explica el comportamiento de los datos, de haber evaluado su idoneidad y de obtener sus predicciones a futuro, hay que preguntarse que hemos aprendido con todo esto y que conclusiones se pueden sacar.

La LEY 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos, establece que los registros electrónicos de la Administración deben de estar disponibles para los ciudadanos las 24 horas del día y los 365 días del año. Además, dicha Ley establece el derecho de los ciudadanos a relacionarse con la Administración por vía electrónica, suprimiendo así el papel. La suma de tener disponibles los trámites y servicios de la Administración a través de Internet y la obligatoriedad de que el registro electrónico esté permanentemente abierto, ha dado como resultado que todos los trámites y servicios estén siempre a disposición del ciudadano por vía electrónica.

Una característica muy acusada de la serie de consultas a la Base de Datos de Ayudas e Incentivos para Empresas, es su estacionalidad semanal, presentando un acusado descenso en las peticiones efectuadas durante el fin de semana. Cabe por ello preguntarse si hay que asignar los recursos y medios durante el fin de semana en igual medida que durante la semana laboral.

Otro rasgo distintivo de la serie es su constante crecimiento. Recordemos que en un período de dos años, partiendo de mínimos de 11 accesos diarios, se han llegado a registrar más de 14.000. Se ha demostrado además que cuando se publicita y se da a conocer el servicio que presta la base de datos, la demanda de la sociedad responde positivamente, señal de su utilidad.

La Administración vive en este momento un período de contracción debido a la crisis que hace que se replantee qué servicios puede dar y en qué medida. Este estudio dota de argumentos a los gestores para defender ante la Dirección General la necesidad de sufragar y, en su caso incrementar, los servidores que alojan la base de datos, adquirir las licencias de software necesarias y no prescindir del equipo de documentalistas que gestionan y mantienen la base de datos y a su vez se encargan de analizar y estructurar la información que en dicha base de datos debe de ser incluida.

Los trabajos de investigación no deben de quedar aislados y adquieren mucho más valor cuando se les dota de continuidad. Con todo lo aprendido hasta ahora, es el momento de preguntarse las cuestiones que se pueden abordar en un futuro. Tal vez en su momento aparezcan otras pero hoy por hoy quedan planteadas la siguientes.

a) **Estudio de Confirmación.**

Este es sin duda el más habitual. Es una repetición del análisis aquí realizado dentro de un período de tiempo; como puede ser un año. Con este estudio se determinará si los datos siguen comportándose de la misma forma o expresado de otra manera, si el modelo aquí obtenido sigue estando vigente. Además se deberá de poder concluir si existen nuevas necesidades de recursos para mantener el servicio de la BD de Ayudas.

b) **Efecto de la Crisis.**

Las crisis económicas son habituales en los países occidentales. Es un fenómeno cíclico que viene a repetirse con un lapso de tiempo de una década entre una y otra. La base de datos de este análisis recoge información útil para la financiación de las empresas, siendo esta un problema típico en época de vacas flacas. Sería útil poder cotejar observaciones recogidas en época de crisis con observaciones recogidas en época de bonanza y poder determinar si la crisis tiene efecto sobre el número de peticiones a la base de datos.

Quedan aquí recogidos estos dos posibles análisis a realizar en un futuro. Por último añadir que en el mundo de la informática el ciclo de vida de los servicios desarrollados, productos y herramientas de software es muy corto. Esto dificulta la continuidad y el seguimiento de cualquier investigación. Pero siempre que se pueda es bueno dotar de continuidad estos estudios.

Bibliografía.

ALONSO REVENGA, JUANA M^a “Apuntes de Series Temporales” 2.012

BISQUERRA ALZINA, RAFAEL “Introducción conceptual al Análisis Multivariante” Editorial PPU

PEÑA SÁNCHEZ DE RIBERA, DANIEL “ESTADISTICA: Modelos y Métodos” Alianza Editorial.

PÉREZ LÓPEZ, CESAR “Econometría Básica. Aplicaciones con EVIEWS, STATA, SAS y SPSS”
Grupo Editorial Garceta 2.012

PÉREZ LÓPEZ, CESAR “Series Temporales: Técnicas y Herramientas.” Grupo Editorial Garceta
2.011

SAS INSTITUTE INC. “SAS/ETS 9.2 User’s Guide”. 2.008



Cuadernos de Trabajo

Facultad de Estudios Estadísticos

-
- CT02/2014** **Values of games with weighted graphs**
E. González-Arangüena, C. Manuel y M. del Pozo
- CT01/2014** **Estimación de la tasa de retorno de la carta del censo de los Estados Unidos a través del modelo de regresión lineal y técnicas de predicción inteligentes.**
José Luis Jiménez-Moro y Javier Portela García Miguel
- CT03/2013** **Provisión de siniestros de incapacidad temporal utilizando análisis de supervivencia.**
Ana Crespo Palacios y Magdalena Ferrán Aranz
- CT02/2013** **Consumer need for touch and Multichannel Purchasing Behaviour.**
R. Manzano, M. Ferrán y D. Gavilán
- CT01/2013** **Un método gráfico de comparación de series históricas en el mercado bursátil.**
Magdalena Ferrán Aranz
- CT03/2012** **Calculando la matriz de covarianzas con la estructura de una red Bayesiana Gaussiana**
Miguel A. Gómez-Villegas y Rosario Susi
- CT02/2012** **What's new and useful about chaos in economic science.**
Andrés Fernández Díaz, Lorenzo Escot and Pilar Grau-Carles
- CT01/2012** **A social capital index**
Enrique González-Arangüena, Anna Khmel'nitskaya, Conrado Manuel, Mónica del Pozo
- CT04/2011** **La metodología del haz de rectas para la comparación de series temporales.**
Magdalena Ferrán Aranz
- CT03/2011** **Game Theory and Centrality in Directed Social Networks**
Mónica del Pozo, Conrado Manuel, Enrique González-Arangüena y Guillermo Owen.
- CT02/2011** **Sondeo de intención de voto en las elecciones a Rector de la Universidad Complutense de Madrid 2011**
L. Escot, E. Ortega Castelló y L. Fernández Franco (coords)
- CT01/2011** **Juegos y Experimentos Didácticos de Estadística y Probabilidad**
G. Cabrera Gómez y M^a.J. Pons Bordería
- CT04/2010** **Medio siglo de estadísticas en el sector de la construcción residencial**
M. Ferrán Aranz
- CT03/2010** **Sensitivity to hyperprior parameters in Gaussian Bayesian networks.**
M.A. Gómez-Villegas, P. Main, H. Navarro y R. Susi
- CT02/2010** **Las políticas de conciliación de la vida familiar y laboral desde la perspectiva del empleador. Problemas y ventajas para la empresa.**
R. Albert, L. Escot, J.A. Fernández Cornejo y M.T. Palomo
- CT01/2010** **Propiedades exóticas de los determinantes**
Venancio Tomeo Perucha

- CT05/2009** **La predisposición de las estudiantes universitarias de la Comunidad de Madrid a auto-limitarse profesionalmente en el futuro por razones de conciliación**
R. Albert, L. Escot y J.A. Fernández Cornejo
- CT04/2009** **A Probabilistic Position Value**
A. Ghintran, E. González-Arangüena y C. Manuel
- CT03/2009** **Didáctica de la Estadística y la Probabilidad en Secundaria: Experimentos motivadores**
A. Pajares García y V. Tomeo Perucha
- CT02/2009** **La disposición entre los hombres españoles a tomarse el permiso por nacimiento. ¿Influyen en ello las estrategias de conciliación de las empresas?**
L. Escot, J.A. Fernández-Cornejo, C. Lafuente y C. Poza
- CT01/2009** **Perturbing the structure in Gaussian Bayesian networks**
R. Susi, H. Navarro, P. Main y M.A. Gómez-Villegas
- CT09/2008** **Un experimento de campo para analizar la discriminación contra la mujer en los procesos de selección de personal**
L. Escot, J.A. Fernández Cornejo, R. Albert y M.O. Samamed
- CT08/2008** **Laboratorio de Programación. Manual de Mooshak para el alumno**
D. I. de Basilio y Vildósola, M. González Cuñado y C. Pareja Flores
- CT07/2008** **Factores de protección y riesgo de infidelidad en la banca comercial**
J. M^a Santiago Merino
- CT06/2008** **Multinationals and foreign direct investment: Main theoretical strands and empirical effects**
María C. Latorre
- CT05/2008** **On the Asymptotic Distribution of Cook's distance in Logistic Regression Models**
Nirian Martín y and Leandro Pardo
- CT04/2008** **La innovación tecnológica desde el marco del capital intelectual**
Miriam Delgado Verde, José Emilio Navas López, Gregorio Martín de Castro y Pedro López Sáez
- CT03/2008** **Análisis del comportamiento de los indecisos en procesos electorales: propuesta de investigación funcional predictivo-normativa**
J. M^a Santiago Merino
- CT02/2008** **Inaccurate parameters in Gaussian Bayesian networks**
Miguel A. Gómez-Villegas, Paloma Main and Rosario Susi
- CT01/2008** **A Value for Directed Communication Situations.**
E. González-Arangüena, C. Manuel, D. Gómez, R. van den Brink



UNIVERSIDAD COMPLUTENSE
MADRID